

REPROGRAMMING HUMAN BLOOD CELLS

by
Bin-Kuan Chou

A dissertation submitted to Johns Hopkins University in conformity with
the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
November 2013

Abstract

Reprogramming differentiated somatic cells to developmentally more potent states opens up a new era in regenerative medicine and research. In particular, the generation of induced pluripotent stem cells (iPSCs) which are expandable and pluripotent offers unprecedented opportunities for disease modeling, drug screening and autologous cell therapy.

However, the lengthy preparation of fibroblasts as starting materials and the genomic integration of viral vectors associated with the traditional iPSC technology impede its wide applications in basic research and therapy. Therefore, identifying a more accessible source for reprogramming and development of non-virus, non-integrating reprogramming technology is highly desired.

Blood cells offer several advantages compared with other cell types. They are easy to obtain from cord blood and peripheral blood, and can expand in a relative short period. In this thesis, I devised several culture conditions to propagate different populations of blood cells and demonstrated that blood cells are facile for efficient reprogramming. Using the blood cells, I also tested a novel set of episomal vectors that can replicate as extra-chromosomes and express key reprogramming transgenes for iPSC generation. After one round of transfection, both CD34 enriched progenitor cells and the expanded monocuclear cells (MNCs) can be efficiently reprogrammed; numerous bona fide iPSC clones were generated from patients and healthy donors. Detailed studies showed these iPSCs were free of episomal vector integration and maintained pluripotency.

Complementary to the iPSC generation, I sought to reprogram human blood cells to clinically more relevant cell types. Lifelong blood cell generation is dependent on rare hematopoietic stem/progenitor cells (HSPCs) to self-renew and generate all blood lineages. HSPCs hold promise in clinical transplantation but *ex vivo* expansion of HSPCs is still largely unsuccessful. Based on recent progress in direct reprogramming, I hypothesized that HSPCs can be generated from differentiated blood cells. I manipulated lineage-committed blood cells with standard reprogramming factors and cultured the cells in conditions that favor HSPC expansion. Strikingly, a group of cells expressing HSPC markers emerged after reprogramming. *In vitro* characterization further proved these cells can differentiate into multiple blood lineages. This study has potentially wide-ranging implications for the direct derivation of HSPCs from blood cells.

Thesis committee

Linzhao Cheng, Ph.D. (advisor and reader)

Zack Wang, Ph.D. (reader)

Roger Reeves, Ph.D.

William Matsui, MD.

Acknowledgements

One of the joys of completion is to reflect on the journey past and remember all the people who have helped and supported me along this long but fulfilling road.

First, I would like to gratefully and sincerely thank my mentor -- Dr. Linzhao Cheng for his inspirational guidance, caring and most importantly, his support during my Ph.D. study. He encourages me to not only grow as an experimental researcher but also as an explorer and an independent thinker. To me, Dr. Cheng is an academic instructor as well as a mind mentor.

I would also like to thank my thesis committee members, professors Roger Reeves, Zack Wang and William Matsui who provided me insightful comments and constructive feedback. I am also truly indebted to Drs. Zhaohui Ye, Jizhong Zou, Prashant Mali and Xiaosong Huang for their guidance and discussion in stem cell biology, genome engineering and hematology in the past years.

Special thanks go to Sarah Dowey for her assistance on experiment and administrative stuff. I would like to thank Kitman Tsang, Cory Smith, Ying Wang, Yongxing Gao, Chaoxia He, Jing Wang, Wei Yen, Sid Shah, Cyndi Liu, and all of the members of the Cheng laboratory. In addition, this dissertation would not have been possible without the help of Dr. Rajini Rao, Colleen Graham, and Leslie Lichter from the Cellular and Molecular Medicine Graduate Program. I am truly indebted and thankful to my friends in Baltimore for their company and sharing. They made my life here colorful and unforgettable. Finally I would like to express my deepest thanks to my family for their unconditional love and support.

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of Figures	x
Chapter 1. Introduction	1
Development and Cell Differentiation	1
Reprogramming Cell Fate.....	2
Current Hurdles in iPSC Generation.....	4
Hematopoietic Stem Cells for Clinical Applications and Remaining Obstacles..	5
Direct Reprogramming – the Concept of Dedifferentiation.....	7
Thesis Aims.....	8
Chapter 2. Human iPSC Derivation from Blood Cells by Non-Integrating	
Plasmids	16
Background.....	16
Approach.....	20
Results.....	21
Expanding Cord Blood CD34 Positive Population in Serum-Free Media.....	21
Reprogramming Human Cord Blood CD34+ Cells by Episomal Vectors.....	22
Reprogramming Human Cord Blood CD34+ Cells by C5 with p53shRNA	

or NANOG.....	23
Sodium Butyrate Can Enhance Reprogramming Efficiency.....	24
Characterization of iPSCs from Cord Blood CD34+ Cells by Episomal Vectors.....	25
Reprogramming Adult Hematopoietic CD34+ Cells by 1-2 Episomal Vectors.....	26
Reprogramming Unfractionated Cord Blood Mononuclear Cells by Episomal Vectors after Stimulation by Cell Culture.....	27
Reprogramming Unfractionated Adult Blood Mononuclear Cells by Episomal Vectors after Stimulation by Cell Culture.....	28
Patient-Specific iPSC Derivation from Unfractionated Peripheral Blood Mononuclear Cells.....	29
PCR Analysis to Confirm the Authentic Origins of iPSC Lines.....	30
Discussion.....	30

Chapter 3. Functional Analysis of the iPSCs Derived from Human Blood

Cells by Non-Integrating Plasmids.....	72
Background.....	72
Approach.....	73
Results.....	74
PCR to Detect the Existence of Episomal Vectors in Derived iPSCs.....	74
Southern Blot Demonstrated the Absence of the Episomal Vectors.....	75

QPCR to Detect the Dynamic Change of Episomal Plasmid Numbers per iPSC at Different Passages.....	75
Whole Genome Sequencing Revealed Absence of Episomal Vector Sequence in the Genomes of the iPSC Lines.....	76
Analysis of DNA Sequence Variations in Derived iPSCs in Comparison with Parental Somatic Cells.....	78
SNVs in Known Functional Elements of the Genome.....	78
Blood Lineage Differentiation of iPSCs Generated from Human Blood Cells by Non-Integrating Plasmids.....	79
Non-blood Differentiation of iPSCs Generated from Human Blood Cells by Non-Integrating Plasmids.....	80
Discussion.....	81

Chapter 4. Engineering Human Induced Pluripotent Stem Cells for GFP

and Hematopoietic-Specific Reporter Systems.....	104
Background.....	104
Approach.....	106
Results.....	107
Generation of Stable GFP Expression iPSC Line by <i>piggyBac</i> Transposon System.....	107
Generate RUNX1-Enhancer/WAS-Promoter Driven <i>piggyBac</i> Vector as the Hematopoietic-Specific Reporter.....	108
Specific Marking of iPSC-Derived Hematopoietic Lineage by the	

RUNX1-Enhancer/WAS-Promoter Driven <i>piggyBac</i> Vector.....	109
Discussion.....	110
 Chapter 5. Direct Reprogramming of Human Lineage Committed Blood	
Cells to Hematopoietic Stem/Progenitor Cells.....	118
Background.....	118
Approach.....	121
Results.....	122
Characterization of Human Cord Blood Mononuclear Cells after	
Erythroblast/Megakaryocyte Expansion.....	122
Generation of Cells with HSPC Traits via Retroviral Vector-Mediated	
Dedifferentiation.....	124
Discussion.....	126
 Chapter 6. Conclusions and Future Perspectives.....	144
Future Perspectives.....	150
 Materials and Methods.....	154
References.....	168
Supplemental References.....	167
Curriculum Vitae.....	188

List of Tables

Table 3-1. List of characterized integration-free human iPSC lines from blood cells by 1-2 plasmids.....	88
Table 3-2. QPCR results of plasmid copy number per iPSC.....	91
Table 3-3. Summary of sequencing two pairs of iPSC lines and their parental somatic cells.....	93
Table 3-4. Sequence variants in the coding regions in BC1 and BCT1 iPSCs.....	94

List of Figures

Chapter 1

Figure 1-1. Waddington's epigenetic landscape to illustrate the biological development and cell fate conversion	11
Figure 1-2. The concept of iPSCs for regenerative medicine applications and current hurdles in iPSC generation	12
Figure 1-3. The human hematopoietic system.....	14

Chapter 2

Figure 2-1. Global epigenetic signatures and relatedness of human ESC lines, iPSC lines and their parental somatic cells.....	33
Figure 2-2. Characterization of expanded cord blood CD34+ cells.....	35
Figure 2-3. Diagrams of novel episomal vectors used in this study, based on an EBNA1/OriP-containing plasmid.....	36
Figure 2-4. Morphology of iPSC-like colonies.....	38
Figure 2-5. Comparison of synergy of pEB-C5 and pEB-Tg episomal plasmids with the combination #6 for cord blood (CB) CD34+ cell reprogramming.....	40
Figure 2-6. Reprogramming efficiency of cord blood (CB) CD34+ cells by various combinations of EBNA1/OriP episomal or transient vectors.....	42
Figure 2-7. Characterization of iPSC lines C7 and CN1 derived from CB CD34+ cells by the single pEB-C5 plasmid.....	44
Figure 2-8. Pluripotency staining of iPSC line C7 and <i>in vitro</i> and <i>in vivo</i>	

differentiation assay.....	46
Figure 2-9. Characterization of an additional iPSC line CTN4 derived from CB CD34+ cells by pEB-C5 and pEB-Tg plasmids in the presence of sodium butyrate (NaB)	48
Figure 2-10. Pluripotency assay of iPSC line CTN4 derived from CB CD34+ cells reprogrammed by pEB-C5 and pEB-Tg plasmids in the presence of sodium butyrate (NaB).....	50
Figure 2-11. Reprogramming of human adult hematopoietic CD34+ cells.....	52
Figure 2-12. Characterization of iPSC line BC1 derived from bone marrow CD34+ cells by the single pEB-C5 plasmid.....	54
Figure 2-13. Pluripotency assay of iPSC line BC1 derived from bone marrow CD34+ cells by the single pEB-C5 plasmid.....	56
Figure 2-14. Growth curves and phenotypes of expanded mononuclear cells (MNCs) from cord blood (CB) or adult peripheral blood (PB).....	58
Figure 2-15. Phenotypes of expanded mononuclear cells (MNCs) from newborn cord blood.....	60
Figure 2-16. Human iPSC derived from cord blood (CB) mononuclear cells (MNCs) by various EBNA1/OriP plasmids.....	62
Figure 2-17. Growth curves and phenotypes of expanded mononuclear cells (MNCs) from adult peripheral blood (PB) and resulting iPSC line characterization.....	64
Figure 2-18. Characterization of disease-specific iPSC line from a HIF2 α mutant patient by retroviral vectors.....	66

Figure 2-19. Pluripotency assay of iPSC line derived from HIF2 α mutant PB CD34+ cells reprogrammed by retroviral vectors in the presence of sodium butyrate (NaB).....	68
Figure 2-20. DNA fingerprint to confirm the authenticity of iPSCs as compared to their somatic cell origins.....	70

Chapter 3

Figure 3-1. Detection of episomal vector sequence by PCR.....	84
Figure 3-2. Southern blot analyses for the lack of vector DNA in expanded iPSCs that are derived by episomal vectors.....	86
Figure 3-3. QPCR to detect the dynamic change of episomal plasmid numbers in reprogrammed cells.....	89
Figure 3-4. Relationship of iPSC lines and their parental somatic cells used in this study.....	92
Figure 3-5. Morphology and phenotypes of BC1 iPSC derived hematopoietic cells.....	95
Figure 3-6. CFC potential of BC1 iPSC derived hematopoietic cells.....	96
Figure 3-7. Erythroblast and megakaryocyte differentiation of BC1iPSCs.....	98
Figure 3-8. Mesenchymal stem cell differentiation of BC1 iPSCs.....	99
Figure 3-9. BC1 iPSCs can differentiate to Schwann cells and PNS neurons.....	101
Figure 3-10. Differentiation of BC1 iPSCs into astrocytes.....	103

Chapter 4

Figure 4-1. Genetic labeling of BC1 iPSCs with GFP via <i>piggyBac</i> transposon vector nucleofection.....	112
Figure 4-2. Expression of WAS-promoter driven <i>piggyBac</i> in blood cells.....	113
Figure 4-3. Specific expression of GFP in BC1 iPSC derived hematopoietic cells.....	115
Figure 4-4. Hematopoietic differentiation induces GFP expression in RUNX1-WAS and WAS nucleofected BC1 reporter iPSCs.....	116
Chapter 5	
Figure 5-1. A model for direct reprogramming to alternative cell fates.....	130
Figure 5-2. A scheme to show the hypothesis that lineage committed blood cells can be directly reprogrammed to more primitive blood cell types but not to pluripotency.....	131
Figure 5-3. Human cord blood mononuclear cells (MNCs) under erythroid expansion culture condition.....	132
Figure 5-4. Characterization of expanded cells.....	133
Figure 5-5. Selection of lineage committed blood cells for the direct reprogramming assay.....	135
Figure 5-6. Infection and culture of reprogrammed blood cells.....	137
Figure 5-7. Lineage committed mononuclear cells successfully infected by the retroviral vector expressed GFP.....	138
Figure 5-8. FACS analysis of reprogrammed cells revealed a emerged CD34+ and CD45+ population.....	139

Figure 5-9. Reprogrammed CD34+CD45+ cells showed multi-lineage differentiation.....	141
Figure 5-10. Differentiated colony-forming cells harvested from the CFC assay showed various blood cell types in the cytopsin assay.....	143
Chapter 6	
Figure 6-1. The paradigm of reprogramming cells for regenerative medicine.....	152

Chapter 1

Introduction

Development and Cell Differentiation

Biological development is a dynamic process of cell growth and differentiation. It begins with versatile stem cells which proliferate and undergo a unidirectional course of differentiation that results in the acquisition of lineage specificity, cellular identity, unique functionality and loss of developmental potential. These differentiated cells become the basic units in the composition of multicellular organisms.

This process was first depicted by Waddington (1957) in his epigenetic landscape more than half a century ago (**Figure 1-1**). Cell fate is determined like a marble rolls down from the top to the lower stable state. Taking a developing embryo (fertilized zygote) as an example, we can imagine it has a status corresponding to a marble at the crest of hill due to its highest potential to give rise to every cell present in the organism and generate an entire animal (**Figure 1-1**). The term “totipotent” is given to describe an embryo’s ultimate differentiation capability. When an embryo starts to grow and divide, the progeny cells gradually differentiate and get distinct fates just like balls rolling down to

different valleys and reaching the point of lowest local elevation. The more a ball falls, the more a cell's program gets fixed and loses differentiation potential.

Reprogramming Cell Fate

It was originally thought such downhill cellular differentiation was an irreversible process during normal development. Cells lost their plasticity forever once they fully differentiated. However, the seminal studies of somatic cell nuclear transfer (SCNT) into enucleated oocytes led by Briggs, King, Gurdon and Wilmut shed the light that cell fate is not unalterable after differentiation (Briggs and King, 1952; Gurdon et al., 1958; Wilmut et al., 1997). The cell fusion of pluripotent cells with differentiated cells provided additional evidence to show that the somatic epigenome can be reset (Cowan et al, 2005). From that time on, scientists realized differentiated cells didn't bear irreversible nuclear changes and still kept all required genetic material capable of making an entire organism.

In 2006, Yamanaka and his group made another big leap by showing that overexpression of four transcription factors can reprogram somatic cells back to a pluripotent state (Takahashi and Yamanaka, 2006). They used retro-viruses to delivery four key genes, Oct4, Sox2, Klf4 and c-Myc into mouse fibroblasts. After selection under embryonic stem cell (ESC) culture conditions, a cell type indistinguishable from ESCs in terms of morphology and property showed up. Yamanaka's group termed these new cells induced pluripotent stem cells (iPSCs) (**Figure 1-2**). Like ESCs, iPSCs can self-renew and harbor the potential to differentiate into cell types of all three germ layers. Their pluripotency was tested

using mouse tetraploid complementation assays and showed mouse iPSCs can grow into an entire mouse (Boland et al., 2009; Kang et al., 2009). In Waddington's landscape, this cellular reprogramming can be portrayed to lift a bottom marble to a summit state (**Figure 1-1**, purple outlined arrow).

The generation of iPSCs offers unique ethical and biological advantages to ESCs in many aspects. First, iPSCs derivation does not involve destruction of human embryos, circumventing many potential moral issues. This makes cellular reprogramming a more attractive method to scientists and policy makers than nuclear reprogramming, which requires sacrifice of human embryos. In addition, cells used for reprogramming can be taken directly from patients. The resulting iPSCs are genetically identical to the original donors. Thus, cells and tissues derived from iPSCs transplanted to patients would not result in immune rejection. Although more studies are required to prove iPSCs do not elicit immune responses, using a patient's own sample for reprogramming lends credit to future autologous cell transplant. It may also eliminate the need for immunosuppressive treatments.

iPSCs also have unprecedented value for disease modeling and drug development (**Figure 1-2**). Many iPSC lines carrying different disease-related mutations have been reported thus far (reviewed by Cherry and Daley, 2013). Patient-derived iPSCs maintain the disease genotype and can be used to study the pathogenic phenotypes and screen drugs to cure diseases. All of these great discoveries made from nuclear and cellular reprogramming became important foundations for stem cell biology and regenerative medicine.

Current Hurdles in iPSC Generation

The capability to restore pluripotency from somatic cells provides hope for cell therapies. However, before iPSCs can be widely considered for scientific research and clinical treatments, several issues should be addressed, including cell sources for iPSC derivation, reprogramming efficiency and reprogramming methods. First, iPSC derivation commonly utilizes skin fibroblasts, which involves local anesthesia for donor skin biopsy (**Figure 1-2**). Such process is labor-intensive and takes time to grow to enough cells before reprogramming can be conducted. In addition, human skin fibroblasts may have been exposed to ultraviolet irradiation and have increased risk of harboring genetic abnormalities. Sampling of blood is one of the least invasive but most routinely performed clinical procedures. A sufficient amount of blood cells can be harvested and expanded in short period of time. Furthermore, blood cells that are derived from hematopoietic stem cells (HSCs) in marrow before being released into circulation may be less exposed to environmental mutagens. Besides, blood such as cord blood with different human leukocyte antigen haplotypes has been collected by blood banks, which can become a source for histo-compatible iPSC generation. These advantages make blood an ideal cell source for iPSC derivation.

Although reprogramming efficiency in mice has dramatically increased over the last few years, with some studies even showing >10% efficiency (Anokye-Danso et al., 2011), the successful rate of reprogramming in human cells remains poor. The efficiency is usually below 0.1% (Ye et al., 2012). The

efficiency and fidelity of reprogramming needs to be improved in order to get sufficient iPSC clones from potential rare samples. In addition, initial methods used to generate iPSCs were viral vector-mediated (**Figure 1-2**). Both viral sequence and transgenes can permanently integrate into the genome, leading to potential insertional mutations and disruption of genome integrity. Besides, some of the transgenes used for reprogramming, like c-Myc and Klf4, are oncogenes and have been documented as highly related to cancer initiation (Albiñ et al., 2010; Tian et al., 2010; Rowland and Peeper, 2006). Thus, creating iPSC lines that are virus-free and with no transgene integration is a goal I want to achieve.

Hematopoietic Stem Cells for Clinical Application and Remaining Obstacles

The hematopoietic hierarchy has been considered the most well studied stem cell system in human body. It includes the most primitive blood cells called hematopoietic stem cells (HSCs) residing at the top of hierarchy (**Figure 1-3**). HSCs can self-renew indefinitely and have the potential to generate every lineage of blood cells found in the hematopoietic system. Everyday a portion of HSCs exit the self-renewing circuit and differentiate into multi-potent progenitors committed progenitor cells which are shown in the middle part (**Figure 1-3**). Following the stream of differentiation, they finally give birth to mature blood cells that locate at the end of the hierarchy (**Figure 1-3**, lower portion). Mature cells are terminally differentiated cells with specific functions and consist of different cell types, including red cells (erythrocytes), white cells, platelets, lymphoid cells,

etc. The whole process is tightly regulated and any tiny flaw may cause a detrimental effect or even threaten the growth of the entire organism.

To date, HSC transplantation is still the only available and routinely performed stem cell-based therapy in clinic. This is because HSCs have the ability to reconstitute the whole blood system once transplanted to the recipient. Therefore, derivation and expansion of these valuable cells efficiently would be a big asset to the clinical hematology.

Attempts to expand HSCs *ex vivo* have been exhaustively investigated for many decades. Approaches covering from signaling pathway studies to bioactive compound screening have been tested, but no satisfactory results have been achieved yet (Blank et al., 2008; Staal and Luis, 2010; Trowbridge et al., 2006; Boitano et al., 2010; Takizawa, et al., 2011; Hofmeister et al., 2007). The expanded cells usually do not fit the qualitative and quantitative requirement for transplantation. The difficulty of obtaining HLA-matched HSCs for allogeneic HSC transplantation also complicates the strategy of *ex vivo* expansion.

The generation of pluripotent stem cells (PSCs) like ESCs and iPSCs opens a new platform to generate HSCs. PSCs are self-renewing cells and can be cultured without losing their pluripotency features. Many *in vitro* differentiation systems have been developed, either mediated by aggregation of PSCs into three-dimensional spheres called embryoid bodies (EBs) or by two-dimensional co-culture with stromal cells. However, the *in vitro* derived hematopoietic cells from PSCs still lack robust *in vivo* engraftment and repopulation activities in conditioned recipient mice. Taken together, it remains a large challenge to ex

vivo expand HSCs or differentiate PSCs to HSCs. One of the reasons is that we cannot recapitulate the microenvironment necessary for normal blood development and HSC maintenance. In addition, not all the cell types and interactions supporting HSC proliferation have been identified. Thus, a novel strategy toward generating transplantable HSCs is highly desired.

Direct Reprogramming – the Concept of Dedifferentiation

Many different kinds of animals show a certain level of regeneration capability to produce new cells, tissues even organs to repair old, damaged or lost ones (Brockes and Kumar, 2002). Two of the most well-known examples would be the salamander and star fish. They both can regrow a new limb when the original one is cut off. In planarian, this feat can be more amazing -- a single one can be cut into hundreds of pieces and each will grow back into a whole planarian.

However, the mechanism behind natural regeneration remains poorly understood. It is speculated that cellular dedifferentiation could be involved in this process. Dedifferentiation implies that terminally differentiated cells are turning back to a more primitive status within their original lineage. By analogy, we can imagine that dedifferentiation is like pushing a marble uphill to a higher but less stable equilibrium (**Figure 1-1**, red outlined arrow). This allows cells have chance to replicate again before redifferentiating to substitute lost cells.

Somatic cell nuclear transfer (nuclear reprogramming) and induced pluripotency (cellular reprogramming) both can be seen as ultimate forms of

dedifferentiation (**Figure 1-1**, purple outlined arrow). They dedifferentiate a somatic cell, which lies on the bottom of epigenetic landscape (foot of hill), back to the very beginning of pluripotent state (summit). By studying the cellular and molecular mechanisms that regulate dedifferentiation and maintain stem cells, we may use a dedifferentiation approach to generate valuable stem cells such as HSCs from differentiated cells to repair damaged cells and tissues one day.

My Aims

Based on current challenges in iPSC technology and hematopoietic stem cell expansion, my goals for the Ph.D. study will be to reprogram somatic cells to iPSCs in a safe and efficient manner. I will find feasible cell types as the starting material and use efficient techniques to generate high quality iPSCs. In addition, I will reprogram differentiated cells to clinically relevant stem cells such as HSCs. If successful, this work will broaden our knowledge of reprogramming and bring stem cells to clinic a step closer.

Specific Aim 1: Human iPSC Derivation from Blood Cells by Non-Integrating Plasmids

Blood cells are easy to derive and have closer epigenetic signatures to ESCs than age-matched fibroblasts, thus I will use blood cells as the starting material. I will determine the efficacy of novel sets of EBNA1/OriP episomal vectors encoding crucial reprogramming factors to derive iPSCs. In the beginning, CD34⁺ cells that represent progenitors from cord blood will be reprogrammed,

and followed by adult bone marrow and peripheral blood CD34+ cells. I will also devise culture conditions to expand unfractionated MNCs for reprogramming. In aim 1, I will also investigate to derive patient-specific iPSCs from patient's blood samples.

Specific Aim 2: Functional Analysis of the iPSCs Derived from Human Blood Cells by Non-Integrating Plasmids

In addition to standard characterization of iPSCs, it is very important to prove that the non-integrating technique won't result in insertion or genome alteration. Several assays will be performed, including PCR and whole genome sequencing. It would also be interesting to test the differentiation potential of blood-derived iPSCs by non-integrating plasmids.

Specific Aim 3: Engineering Human Induced Pluripotent Stem Cells for GFP and Hematopoietic Specific Reporter Systems

Candidates of hematopoietic specific promoters/enhancers will be cloned to regulate GFP expression. By testing the reporter system in human iPSCs, it may serve as a good platform to improve hematopoietic differentiation in the future.

Specific Aim 4: Direct Reprogramming of Human lineage Committed Blood Cells to Hematopoietic Stem/Progenitor Cells (HSPCs)

Recent studies of induced pluripotency (Buganim et al., 2013) and direct cell conversion (Ladewig et al., 2013) (**Figure 1-1**, blue outlined arrow) reshaped our view of cellular identity and plasticity. These findings inspired me to ask if we can generate clinically meaningful cells such as HSPCs via cellular reprogramming. Based on this notion, I would like to directly reprogram lineage committed blood cells not all the way back to the pluripotent state, but to guide them to a hematopoietic primitive state. The standard reprogramming factors will be used to reprogram human lineage committed blood cells, and reprogrammed cells are going to be cultured in optimized HSPC culture condition for potential HSPC generation. This process can also be illustrated as “dedifferentiation” (**Figure 1-1**, red outlined arrow).

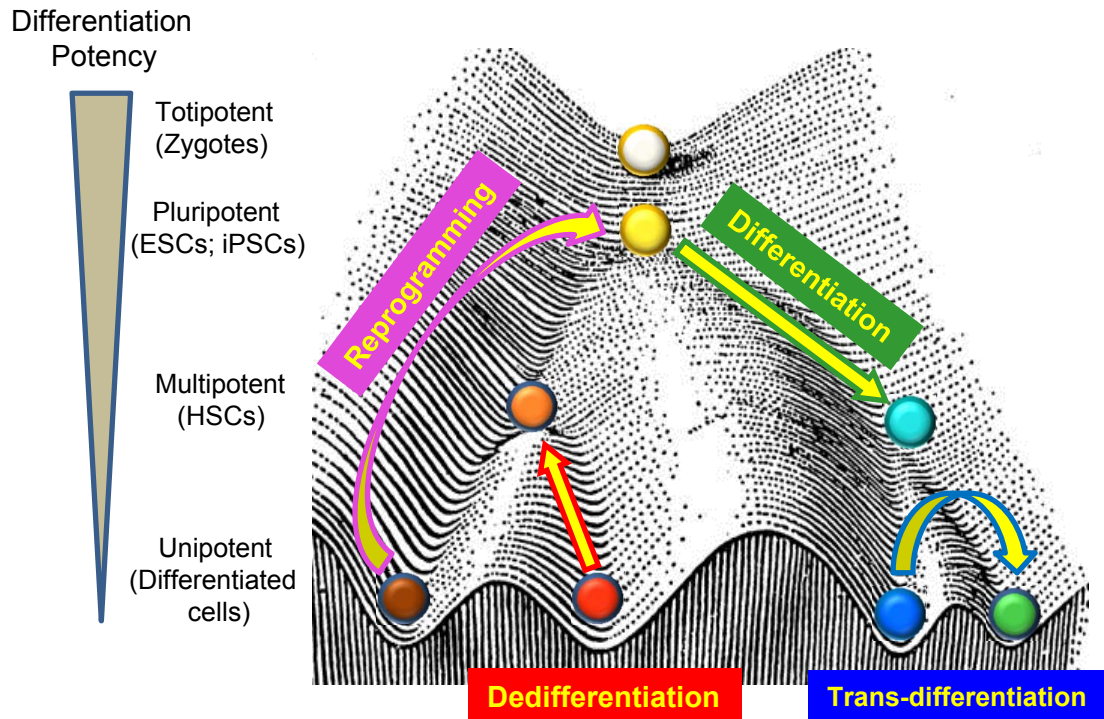


Figure 1-1. Waddington's epigenetic landscape to illustrate the biological development and cell fate conversion. Differentiating cells are depicted as a ball rolling down a hill. Each concave represents a stable cell lineage. The height of a cell positively reflects the differentiation potential. Arrows are used to depict how cells can convert to other cell fates. The figure is reproduced from the original text by Waddington. Abbreviations: ESCs, embryonic stem cells; iPSCs, induced pluripotent stem cells; HSCs, hematopoietic stem cells.

- Lengthy preparation of fibroblasts
- No easy accessible source
- Low reprogramming efficiency
- Insertional oncogenic risk
- Disruption of genome integrity

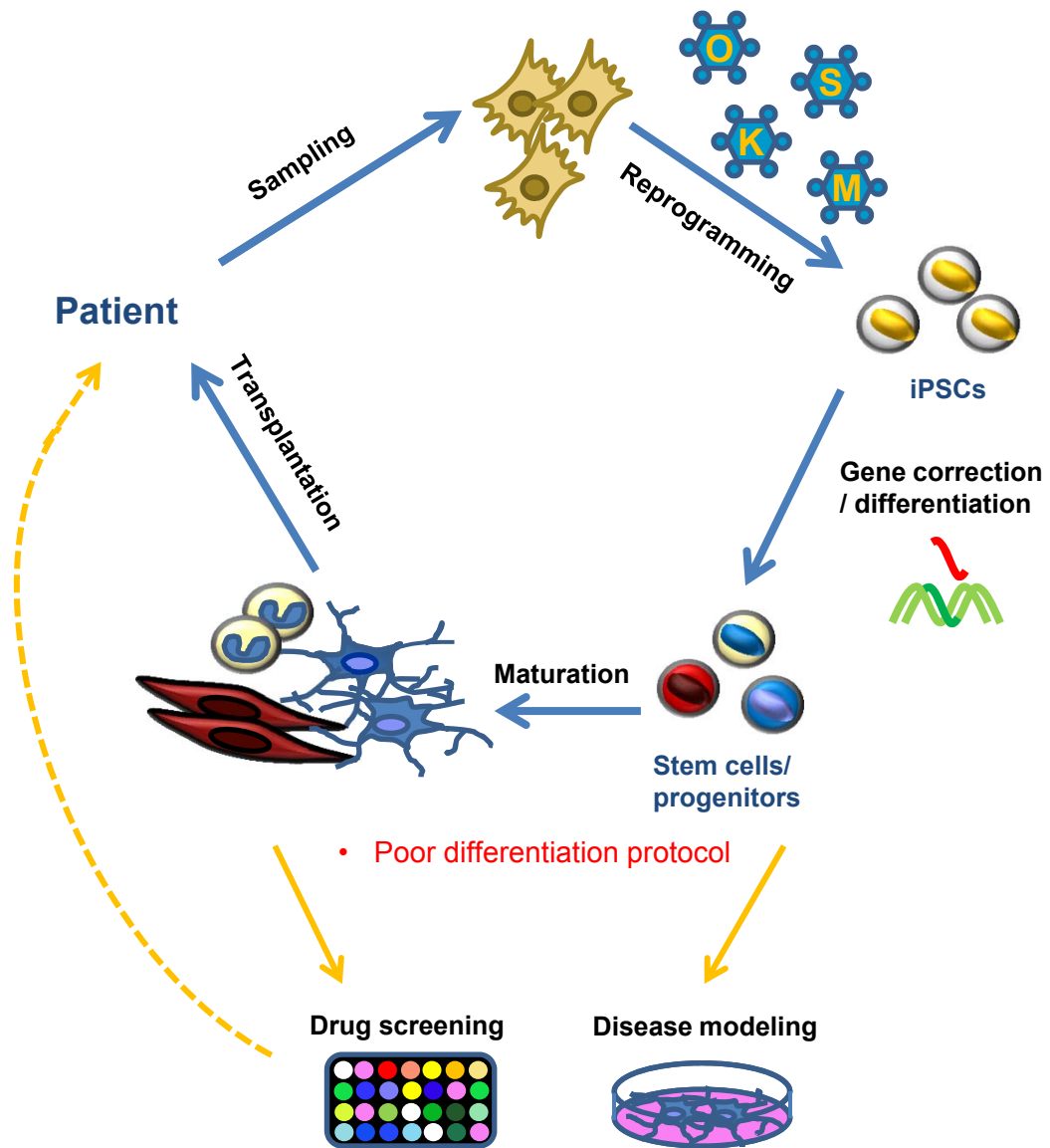


Figure 1-2. The concept of iPSCs for regenerative medicine applications and current hurdles in iPSC generation. iPSCs from patients can be used for disease modeling, drug screening and potential autologous cell therapy.

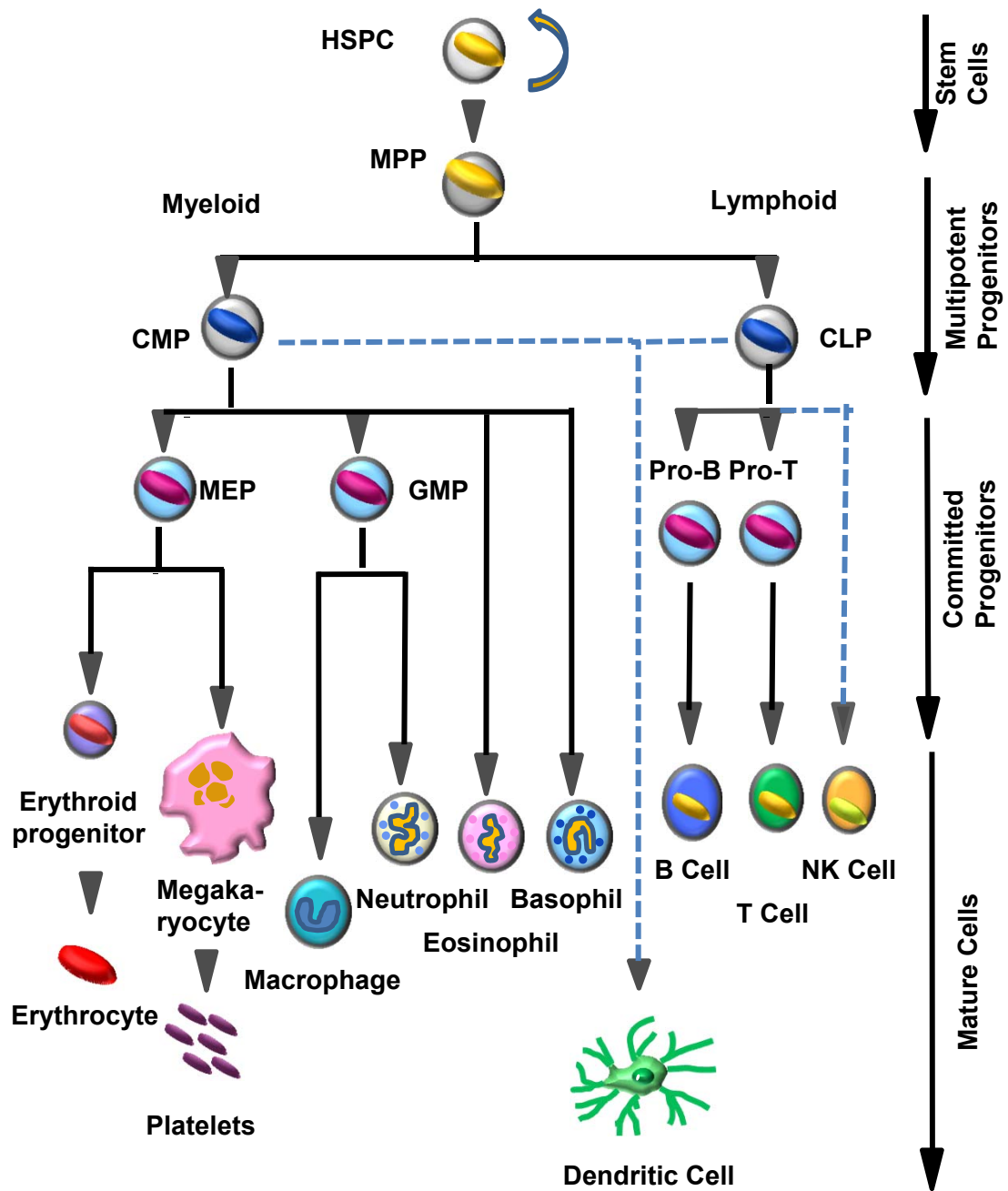


Figure 1-3. The human hematopoietic system. Hematopoietic stem cells reside at the top of hierarchy which can self-renew and give rise to multiple lineages of blood progenitors and mature blood cells. Abbreviations: MPP, multi-potent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, Megakaryocyte /erythrocyte progenitor; GMP, granulocyte/macrophage progenitor.

Chapter 2

Human iPSC Derivation from Blood Cells by Non-Integrating Plasmids

Background

Human induced pluripotent stem cells (iPSCs) that morphologically and functionally resemble human embryonic stem cells (ESCs) have been generated from a wide spectrum of somatic cell types by viral vectors expressing defined reprogramming factors since 2007 (Takahashi et al., 2007; Yu et al., 2007; Park et al., 2008). Fibroblastic cells from a skin biopsy are one of most common sources for iPSC generation. But the lengthy and cumbersome culture prior to reprogramming invoked researchers to find a better source for reprogramming (**Figure 1-2**).

To identify accessible and permissive human cell types for efficient derivation of iPSCs, our lab has investigated epigenetic and gene expression signatures of multiple postnatal tissue types. Our analysis revealed that blood cells display unique signatures that are closer to iPSCs and ESCs than age-matched fibroblastic cells to iPSCs/ESCs, making blood an attractive choice for iPSC derivation (**Figure 2-1**)(Chou et al., 2011). Indeed, blood sampling can

acquire sufficient cell number instantly and has been routinely performed in clinical settings.

Of many blood cell types that can be potential candidates for iPSC generation, cord blood CD34+ cells have several advantages than other blood cells. Cord blood has an abundance of CD34+ cells, which are enriched for hematopoietic stem cells (HSCs) and have higher proliferation potential than CD34+ cells from other blood origins. Cord blood cells also have been widely collected in hospitals and stored in many blood banks and can be acquired. The huge reservoir of cord blood is suitable for basic research or future therapeutic usage. In addition, blood cells from newborn babies have limited, if not nil, exposure to environmental mutagens and minimize the risk of genetic damage. All of these reasons make cord blood an excellent cell type for reprogramming.

CD34+ cells isolated from adult bone marrow which has been investigated extensively over many years is one of the most common methods to purify HSPCs from human (reviewed by Doulatov et al., 2012). In recent years, mobilized peripheral blood (mPB) has become another major source of adult CD34+ cells. This is owing to the discovery that peripheral blood of mice treated with repeated G-CSF drastically increased the colony-forming unit in spleen pool (Molineux et al., 1990). Besides, collecting mPB can bypass the invasive procedure to harvest bone marrow cells. In clinical transplantation, mobilized peripheral CD34+ blood cells have a trend to replace bone marrow CD34+ cells. And one unique advantage of BM and mPB compared to CB is that we can harvest BM or mPB cells from patients who have genetic mutations. The iPSCs

derived directly from patients provide a valuable tool not only in disease modeling, but also in drug discovery, toxicology and even stem cell therapies where immune rejection is no longer a concern. Therefore, it is important to reprogram not only cord blood but also many other types of blood cells in an efficient manner.

However, even in cord blood, the CD34⁺ cells are very rare and isolation of the CD34⁺ cells is cumbersome. For certain diseases like aplastic anemia in which patients have a shortage of CD34⁺ blood cells, isolating CD34⁺ cells would be quite challenging. Therefore, several groups are generating human iPSCs from “unfractionated” blood mononuclear cells (MNCs) (Kunisato et al., 2010; Brown et al., 2010; Seki et al., 2010; Loh et al., 2010; Staerk et al., 2010). These studies involved a long cell culture to activate T cells which are most abundant in peripheral (PB) MNCs into a proliferating state. T cells were easily expanded and infected by viral vectors under these culture conditions. But, detailed characterization revealed that all or the vast majority of the derived iPSCs in each of the 5 studies contain somatic V(D)J DNA rearrangements at the T cell receptor (TCR) loci. It is unclear presently, whether the preexisting genomic rearrangement in human T cell-derived iPSCs will affect iPSC differentiation to T and non-T cell lineages (Yamanaka, 2010; Serwold et al., 2007), and whether T cell-derived iPSCs may form T cell lymphomas at an unexpectedly high rate as observed with reprogrammed mouse T cells (Serwold et al., 2010). Therefore, developing a feasible method to skip CD34⁺ cell

purification and expand blood cells to non-T cell lineage would be a great benefit for iPSC derivation.

An equally critical issue in human iPSC generation is to find an appropriate system to introduce reprogramming factors. Methods of iPSC derivation without altering the genome are preferred, ideally by a virus-free approach. Recently, several virus-free and integration-free methods were reported, which generated mouse and human iPSCs by using purified proteins, mRNAs as well as plasmids (Okita et al., 2008; Zhou et al., 2009; Kim et al., 2009; Jia et al., 2010; Warren et al., 2010; Yu et al., 2009; Marchetto et al., 2009). However, most of these published studies using proteins or mRNAs needed serial delivery (up to 17 times) of multiple reprogramming molecules to targeted adherent cells. Two groups have used plasmid-based episomal vectors to derive iPSCs from human neonatal fibroblasts and fetal neural progenitor cells (Yu et al., 2009; Marchetto et al., 2009). This is based on a well-established technology that inclusion of the EBNA1 gene and the OriP DNA sequence from the Epstein-Barr virus enables a plasmid, after one-time DNA transfection, to replicate in many types of primate cells as a circular episome (Lindner and Sugden, 2007). However, the reported efficiency of episomal vector-mediated reprogramming was very low (~3-6 iPSC clones per 10^6 neonatal human fibroblasts) even when the best combination of 3 EBNA1/OriP plasmids expressing 7 genes was used in DNA transfection (Yu et al., 2009). For cell therapy development and disease modeling, it is necessary to develop more efficient methods to generate integration-free and virus-free human iPSCs from adult somatic cells that tend to

be more refractory to reprogramming as compared to fetal/neonatal cells. Therefore, an improved EBNA1/OriP episomal vector system would be required to reprogram adult and neonatal somatic cells efficiently.

Approach

Finding a suitable starting cell population and a non-integrating method will be the key to derive high quality iPSCs. Our lab has previously reported that CB CD34+ cells can be fully reprogrammed by retroviral vectors (Ye et al., 2009). Based on their easy acquisition and high growth potential, I will start from human cord blood CD34+ cells and expand cells in a chemically-defined culture condition, which is feeder cell free and serum free, to generate sufficient proliferating cells. In the same time, inspired by the potential advantage of using the episomal plasmid in previous studies, my colleague and I cloned essential reprogramming genes, Oct4, Sox2, Klf4, c-Myc and Lin28 into a single episomal vector through linkage of distinct self-cleaving 2A peptides (accredited to P. Mali). The vector consists of EBNA1 gene and an OriP DNA sequence both originally from Epstein-Barr virus. EBNA1 protein binding to OriP sequence can boost vector replication and sustain vector as an episomal form during cell division, thus enhance the expression of transgenes. The easy-handling property of episomal-based system encourages me to explore how an efficient episomal vector can reprogram human cord blood CD34+ cells.

I will test if this novel episomal vector can also reprogram other CD34+ cells such as from BM and mPB. If successful, it would set up a solid foundation

to reprogram blood harvested from general populations and patients. Just in case I have to cope with some cell types that are more refractory to reprogramming, I also seek to refine methods that can enhance overall reprogramming efficiency. Factors and small molecules that have been shown to improve reprogramming will be also utilized.

CD34+ cell isolation is a time-consuming process and may have some detrimental effects on donors. Therefore, I will try to reprogram MNCs either without CD34 purification or from CD34 depleted cells. Instead of growing MNCs to lymphoid lineage where resulting somatic rearrangement may constrain iPSCs' capability to a full spectrum differentiation, I will culture cells in a condition that favors erythrocyte/megakaryocyte growth from unfractionated MNCs. Hopefully erythroblast progenitor cells can be stimulated and maintained in a high proliferative rate which is prerequisite for successful reprogramming (Ruiz et al., 2011). Lastly, blood cells from a patient will be also tested to derive patient-specific iPSCs.

Results

Expanding Cord Blood CD34 Positive Population in Serum-Free Media

Cord blood (CB) cells with or without CD34 pre-fractionation were acquired from AllCells, Ltd. CB CD34+ cells after serum-free media (SFM) and supplemental cytokines culture can expand to 3-5 fold. Prolonged culture proved cells were in a rapidly growing state. It has been documented a high proliferate rate is crucial for efficient reprogramming (Ruiz et al., 2011). The majority of cells

maintain CD45⁺ and CD34⁺ markers (data not shown); CD45 is a pan-hematopoietic cell marker while CD34 marker is an important indicator of HSPCs. In addition, colony forming units (CFU) assay demonstrated cells after SFM expansion still possessed the property to form colonies with different types (**Figure 2-2**). Taken together, these data indicated that CB CD34⁺ cells after expansion are in an appropriate condition for efficient reprogramming.

Reprogramming Human Cord Blood CD34⁺ Cells by Episomal Vectors

I next tested a novel set of EBNA1/OriP plasmids constructed for reprogramming cultured CB CD34⁺ cells, which have the most favorable epigenetic/gene expression signatures and proliferative capacity as compared to adult CD34⁺ cells or fibroblasts/MSCs (Chou et al., 2011). In the first EBNA1/OriP plasmid construct (termed pEB-C5), 5 reprogramming factors from mouse origin: Oct4 (also known as pou5f1), Sox2, Klf4, c-Myc and Lin28, are expressed as a single polycistronic unit through self-cleaving peptides linking (**Figure 2-3A**). In the second set of EBNA1/OriP plasmids, SV40 Large T antigen (pEB-Tg) was constructed to test additive effect of reprogramming, while GFP was used as an efficiency control (pEB-GFP) (**Figure 2-3B**). To test the efficacy of novel pEB-C5 and -Tg, I compared them with the combination of 3 vectors used in previous studies (combination #6) (Yu et al., 2009). After expansion (~5-fold) with cytokines for 4 days, CB CD34⁺ cells were transfected once and then cultured on MEF feeders following the standard protocol for deriving human iPSCs (Ye et al., 2009; Mali et al., 2010) (**Figure 2-3C**). Two small molecules,

sodium butyrate (NaB) and valproic acid (VPA) that have been reported to improve reprogramming were added individually. Strikingly, expanded CB CD34+ cells were very amenable to reprogramming after episomal plasmid nucleofection. The emergence of iPSC-like colonies came up as early as day 5-6 post nucleofection (**Figure 2-4A**, arrow) and hundreds of colonies showed up in next 2 weeks.

Previous studies showed that the acquisition of cell surface expression of TRA-1-60/TRA-1-81 is a better marker for monitoring full reprogramming of human somatic cells (Mali et al., 2010; Chan et al., 2009; Lowry et al., 2008). I performed TRA-1-60 live staining of whole cultures at day 14 and numerated both TRA-1-60 positive (+) and negative (-) colonies (**Figure 2-4B**, **Figure 2-5A**, **B**). Live-staining of TRA-1-60 suggested that most of emerged colonies were bona fide iPSC lines (**Figure 2-4B**, **C**). Our 2-plasmid combination (C5+Tg) generated similar numbers of TRA-1-60+ colonies (**Figure 2-5A**), but their percentages among total colonies were much higher than combination #6 (**Figure 2-5B**). Various TRA-1-60 negative colonies generated by combination #6 resemble mere transformed cells rather than fully reprogrammed clones (**Figure 2-4D**, **E**). This data indicates C5-based reprogramming is competitive to combination #6 in terms of efficiency and can be a more reliable method to derive true iPSCs.

Reprogramming Human Cord Blood CD34+ Cells by C5 with p53shRNA or NANOG

I would like to test whether I can replace Tg by either NANOG or other factors. Therefore, a transient vector expressing human NANOG was generated. In another set of EBNA1/OriP constructs, a small hairpin RNA targeting p53 (termed p53shRNA) was cloned to confer p53 interference effect (**Figure 2-3B**). The purpose is to test if p53 silencing or NANOG expression can deliver the same result as Tg expression. I also addressed if Tg can be omitted completely. The stimulatory effect of CB CD34⁺ cell reprogramming by the second vector over pEB-C5 alone was the greatest by Tg followed by p53shRNA, and minimal by NANOG (**Figure 2-6A**). The pEB-C5 plasmid alone was sufficient to generate about 50 TRA-1-60⁺ colonies from 10⁶ nucleofected cells (from approximately 0.2x10⁶ original CD34⁺ cells (**Figure 2-6A**). This data showed Tg is not essential. In the presence of the second plasmid, the efficiency can be enhanced to close roughly 100 TRA-1-60⁺ colonies from 10⁶ nucleofected cells.

Sodium Butyrate Can Enhance Reprogramming Efficiency

I observed that sodium butyrate (NaB), a nutritional supplement and histone deacetylase (HDAC) inhibitor that stimulates reprogramming of human fibroblastic cells (Mali et al., 2010), also consistently enhanced the number and percentage of TRA-1-60⁺ colonies by either vector combination (**Figure 2-5 and 2-6 A, B**). In the presence of NaB at millimolar level during the reprogramming (day 3-14), about 160 TRA-1-60⁺ colonies were generated in C5 alone condition, reflecting more than 3-fold increase. Moreover, the percentage of TRA-1-60⁺ colonies also raised to 50% of the total colonies formed (**Figure 2-6B**). This

result showed the advantage of including NaB as a supplement to boost reprogramming.

Characterization of iPSCs from CB CD34+ Cells by Episomal Vectors

After TRA-1-60 live staining, I picked individual iPSC-like colonies derived by transfection with either pEB-C5 alone (C), pEB-C5 + pEB-Tg (CT) or CT in the presence of NaB (CTN). I examined at least two CB-derived clones from each category by standard assays. Undifferentiated phenotypes of one representative clone, C7, is shown (**Figure 2-7A**), and additional clones CTN4 is shown (**Figure 2-9A**). I found that 5 of 6 clones derived with pEB-C5 + pEB-Tg vectors have a normal karyotype, including C7, CN1, CT5 and CTN4 (**Figure 2-7B** and **Figure 2-9B**). To further analyze iPSCs derived with transient expression of Tg, I used genome-wide SNP analysis to examine the similarity of the CT5 and CTN4 genome to that of primary CB CD34⁺ cells from the same CB donor. The genome-wide data (covering more than 10^6 SNPs) indicated that CT5 and CTN4 iPSCs are essentially identical to CD34⁺ cells of the CB donor and that reprogramming by the episomal vectors did not cause any detectable alterations in the genome (see Materials and Methods). Equally important, CT5 and CTN4 iPSCs were pluripotent as validated by both *in vitro* and *in vivo* differentiation assays (**Figure 2-10**). Similarly, the iPSC clones derived by the single C5 vector from the same donor (CN1) and a different donor (C7) were also pluripotent (**Figure 2-8**).

Reprogramming Adult Hematopoietic CD34+ Cells by 1-2 Episomal Vectors

I also acquired adult BM CD34+ cells from AllCells, and tested with 1 or 2 EBNA1/OriP plasmids (**Figure 2-11**). Cells were cultured under serum-free and feeder cell-free condition. After 4 days of expansion, 10^6 cultured cells were nucleofected by C5 alone, C5 + Tg or C5 + p53shRNA. The iPSC-like TRA-1-60+ colonies emerged between day 9 and day 11, about 3-4 days later than those from CB CD34+ cells. I found the efficiency of reprogramming adult CD34+ cells was 50-fold lower than CB CD34+ cells per 10^6 cells transfected by the same vector sets, although sufficient numbers of iPSC colonies formed (**Figure 2-11A**). Adding a second EBNA1/OriP vector expressing Tg or shRNA targeting p53 increased the efficiency by ~5-fold.

Since BM aspiration is a complicated process, I next tested CD34+ cells from a more accessible adult source--mobilized peripheral blood (mPB). Again, the same 3 sets of episomal plasmid combinations were used. I found in general the mPB efficiency for iPSCs is similar to BM cells. With the help from Tg, the colony numbers were increased. Meanwhile, p53shRNA was close to Tg to derive iPSCs. One clone from BM CD34+ cells (BC1) was further characterized after 5-12 passages. This BC1 clone reprogrammed by the single pEB-C5 vector was picked, expanded and characterized as before. BC1 showed undifferentiated phenotypes (**Figure 2-12A**), had normal karyotype (**Figure 2-12B**), and was pluripotent by *in vitro* (**Figure 2-13A**) and *in vivo* (**Figure 2-13B**) differentiation assays. These data suggested that although in a low efficiency manner, episomal

plasmids still can reprogram adult CD34⁺ cells and generate reasonable iPS colonies.

Reprogramming Unfractionated Cord Blood Mononuclear Cells by Episomal Vectors after Stimulation by Cell Culture

From the view to simplify the process of iPSC derivation, circumvent the need of fractionation or specific cell type purification is necessary. Thus I attempted to reprogram unfractionated blood MNCs without CD34 purification by the improved episomal vectors. Uncultured or cultured MNCs under the CD34⁺ cell culture condition did not survive well before or after nucleofection, resulting in zero iPSC colonies per five million MNCs (data not shown). I tested various culture conditions that would stimulate myeloid-erythroid proliferation from MNCs, and avoided the standard protocols expanding T and/or B cells that contain V(D)J DNA rearrangements. V(D)J DNA rearrangement indicates some nucleotides in the genome have been trimmed and results in altered genome. I chose a condition that stimulates erythroblast proliferation (van den Akker et al., 2010) for the following 2 reasons. 1) I observed net gains of cell numbers after culture: 2.5 fold by day 8 for CB MNCs and 1.5 fold by day 10-14 for PB MNCs, despite a reduction in the first 2-4 days (**Figure 2-14A**). 2) The expanded cells contained very few T or B cells compared to input MNCs (**Figure 2-14B, C**). Nearly all the expanded cells expressed pan-hematopoietic marker CD45 (**Figure 2-15A**), and many expressed CD235a⁺ erythroid phenotypes (**Figure 2-15A**). There were only minimal CD34⁺ cells (**Figure 2-15A**). Expanded cells also

showed erythroblast morphology at day 8-9 (**Figure 2-15B**), and were enriched for colony-forming erythroid progenitor cells (data not shown). The data suggested that the resulting cells after expansion are enriched for highly proliferative erythroblasts as previously reported (van den Akker et al., 2010).

I devised a condition to efficiently transfect expanded cells with favorable profiles. Two million cells were transfected one-time by nucleofection with the same one to three EBNA1/OriP episomal vectors, and followed by identical reprogramming protocol as before (**Figure 2-3C**). The 8-day-cultured CB MNCs after nucleofection can generate TRA-1-60+ iPSC-like colonies very efficiently (**Figure 2-16A**). Although variations of different CB MNC preparations existed, the overall reprogramming efficiency is very high: 460 and 160 TRA-1-60+ colonies formed by day 14 per 10^6 transfected MNCs by the single pEBC5 episomal vector in the presence of NaB. The stimulatory effect of Tg or NaB is minimal for the cultured CB MNCs (**Figure 2-16A**). However, in the presence of NaB, I can always observe higher percentage of TRA-1-60+ colonies over total emerged colonies, indicating a low incidence of transformed cells (**Figure 2-16B**).

Reprogramming Unfractionated Adult Blood Mononuclear Cells by Episomal Vectors after Stimulation by Cell Culture

I also tested to culture adult unfractionated peripheral blood (PB) MNC for erythroid expansion. Again, cell number decreased in the beginning of culture, but gradually to rebound around one week (day 9, **Figure 2-17A**). By day 15 I got 7.2 million cells which was more than initial input cell number. I reprogrammed

expanded cells by C5 alone or together with Tg, all were in the presence of NaB. Although the efficiency is much lower than that of CB MNCs, I still derived a reasonable number of iPSC colonies from 2×10^6 cultured MNCs (**Figure 2-17B**). If Tg was utilized, I can generate more iPSC lines. Similar to adult PB CD34+ cells, PB MNC reprogramming is more dependent than CB MNCs on NaB stimulation. Adding pEB-Tg further increases the efficiency by roughly 3-fold, resulting in 2 to 37 iPSCs per 2×10^6 cultured MNCs. I further characterized the iPSC lines derived from cultured adult PB MNCs. Pluripotency staining of one expanded iPSC line, PE2, was shown (**Figure 2-17C**). Thus, my method is adequate for iPSC derivation from less than 1 ml of PB.

Patient-Specific iPSC Derivation from Unfractionated Peripheral Blood Mononuclear Cells

To test the idea of establishing iPSCs from a patient, I harvested PB MNCs from a patient with HIF2a gain-of-function mutation. In adult mice, HIF-2 α gene plays a key role in regulating erythropoietin production and erythropoiesis (Semenza, 2009). This patient's red blood cells are very sensitive to erythropoietin, indicating this mutation is relevant to hematopoiesis (Percy et al., 2008a; Percy et al., 2008b; Furlow et al., 2009). CD34 positive and negative populations were separated by MACS CD34+ cell kit. The CD34+ population was reprogrammed by standard OSKM retrovirus in early days before episomal plasmid had been developed. **Figure 2-18A** showed the pluripotency staining and normal karyotyping of one iPSC line. The CD34- counterpart was expanded

under our erythroid culture condition. After C5 + Tg reprogramming, I derived average 37 iPS clones from 2 million starting cells. Genomic DNA was extracted and sent for sequencing and confirmed the existence of mutation (**Figure 2-18B**). One HIF2a-mutation iPSC line was characterized, including *in vitro* and *in vivo* differentiation shown in **Figure 2-19**. Therefore, this is proof-of-principle that patient-specific iPSCs can be readily derived using retroviral vector or non-integrating plasmid. The iPSC line carrying the disease genotype would be ideal for disease modeling.

PCR Analysis to Confirm the Authentic Origins of iPSC Lines

To confirm the relationship between the starting cell population and derived iPSCs, genomic DNAs were isolated from various iPSC lines and their parental cells such as cord blood (CB), bone marrow (BM) CD34+ cells and MSCs, and adult peripheral blood (PB) mononuclear cells (MNCs). They were used as templates for detection of variable regions (in size) in the genome by PCR using two sets of specific primers. Based on the identical size of bands and patterns, I confirmed all iPSC lines were derived from corresponding original cells (**Figure 2-20**).

Discussion

The efficient reprogramming is largely dependent on the intrinsic property of starting cells. In this part of research, I found human cord blood CD34+ cells were an easily accessible and amenable cell type for reprogramming. They can

be propagated in a serum- and feeder-free conditions and reach sufficient cell number in a short time of culture for subsequent experiments.

To focus on generating integration-free iPSCs, I employed a novel episomal vector system to transfect briefly-cultured CB CD34+ cells. The first iPSC-like colony can show up as early as day 6 post-transfection, which is much faster than traditional virus-mediated reprogramming. The overall reprogramming efficiency was also high, making it easy to pick up enough clones for detailed studies. Notably, although our novel episomal vector system generated similar iPSC colony numbers as previously reported episomal vectors did (Yu et al., 2009), most emerged colonies in our system were TRA-1-60 positive. In contrast, colonies derived by combo#6 (**Figure 2-5B**) showed a significant portion of TRA-1-60 negative. It suggests that my system is much more efficient to confer full reprogramming and facilitating the efficient isolation of the desired TRA-1-60+ colonies.

In line with our previous study, adding NaB can enhance the efficiency of iPSC derivation for either fibroblasts or blood cells. Adding NaB continuously during the reprogramming did not negatively affect the derivation of iPSCs. This data supports that NaB could be included in the standard procedure.

Transient expression of a second vector (pEB-Tg) or shRNA targeting p53 did not show stimulatory effect, although Tg stimulation and p53 inhibition on reprogramming were observed by several groups using different vectors and somatic cell types (Yu et al., 2009; Mali et al., 2008; Zhao et al., 2008; Krizhanovsky and Lowe, 2009). This could be due to the high reprogramming

efficiency in CB CD34⁺ cells. Transient expression of Tg by episomal vectors did not necessarily alter karyotypes of the derived iPSCs (Yu et al., 2009; Chou et al., 2011), hence pEB-Tg may play a more important role if refractory cell types are going to be reprogrammed. Taken together, my platform is more flexible if one needs to omit or replace Tg with other factors for mechanistic investigations.

This episomal vector system, together with optimized cell culture conditions, allows us to efficiently generate virus-free human iPSCs from blood MNCs isolated from a few milliliters of blood after suspension cultures for 8-9 days. When further enhancement is needed for other cell types, or patient's blood MNCs that are more refractory to reprogramming by the current episomal vectors, I can easily add additional reprogramming genes over the 5-6 transgenes I used. Since I can deliver multiple (at least 8) genes from a single EBNA1/OriP plasmid for simultaneous and transient expression, it is no longer critical to reduce the numbers of reprogramming factors used. Obviously, small numbers of vectors (1-2) are preferred to ensure the terminal loss of episomal vector after reprogramming is achieved.

Distance

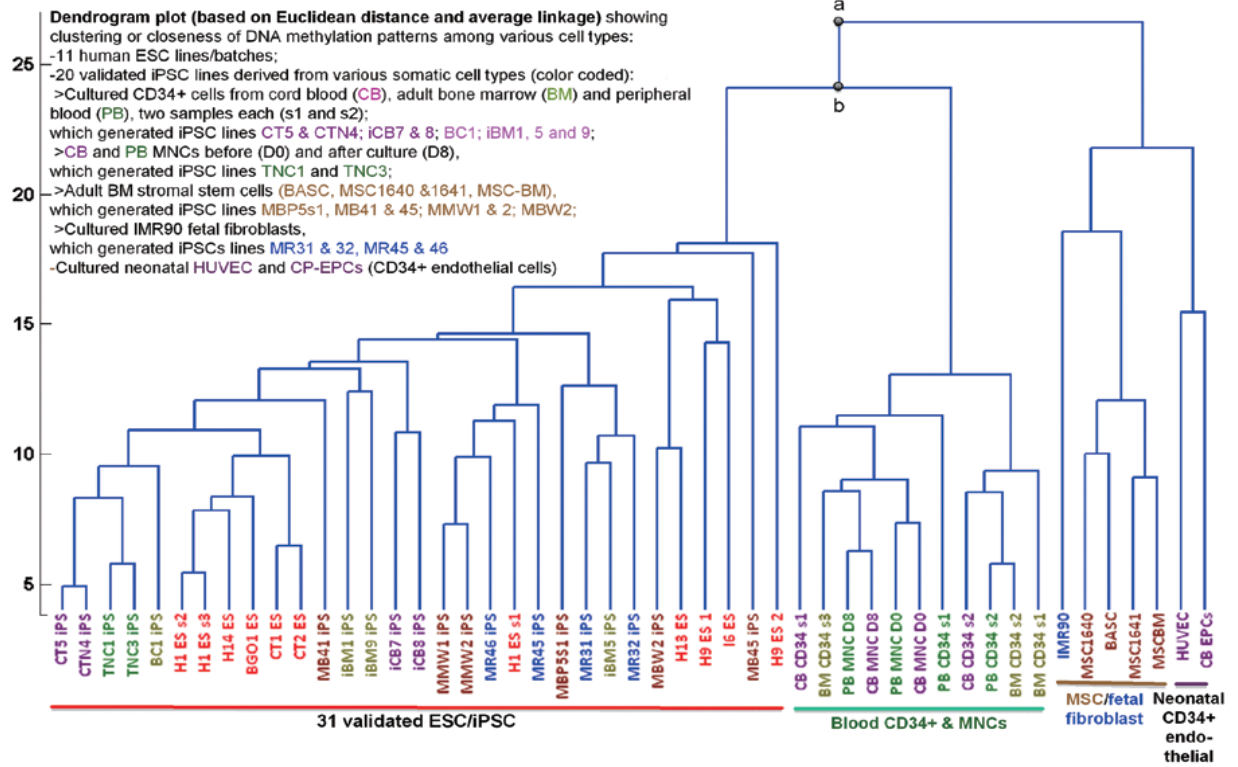


Figure 2-1. Global epigenetic signatures and relatedness of human ESC lines, iPSC lines and their parental somatic cells. A dendrogram plot using Euclidean distances and average linkage based on DNA methylation at 26 424 autosomal loci in the Infinium Methylation27 platform, to analyze promoter DNA methylation genome-wide. 11 hESCs (passages 30-80) and 20 iPSCs derived from various somatic cell types (passage 5-15) that have a normal karyotype and validated pluripotency were analyzed. Parental cell types used for reprogramming include cultured CD34+ cells from cord blood (CB), adult peripheral blood (PB) and adult bone marrow (BM), CB and PB mononuclear cells (MNCs) before (D0) and after culture (D8), 4 types of adult BM stromal cells (MSC-BM, MSCs and BASC) and IMR90 fetal fibroblasts, neonatal cord-derived endothelial cells (CD34+) such as HUVEC and CB-derived endothelial progenitor cells (CB-EPCs). The clustering confirms that the 20 validated iPSC lines derived from various somatic cell sources are highly similar to ESCs, but distinct from their parental cells. DNA methylation patterns in human blood cells are more similar to ESC/iPSC group as compared to age-matched non-hematopoietic cells. Based on a two-sample *t*-test with unequal variances, we confirmed that the different mean distance corresponding to the indicated nodes **a** and **b** is statistically significant ($P < 0.01$).

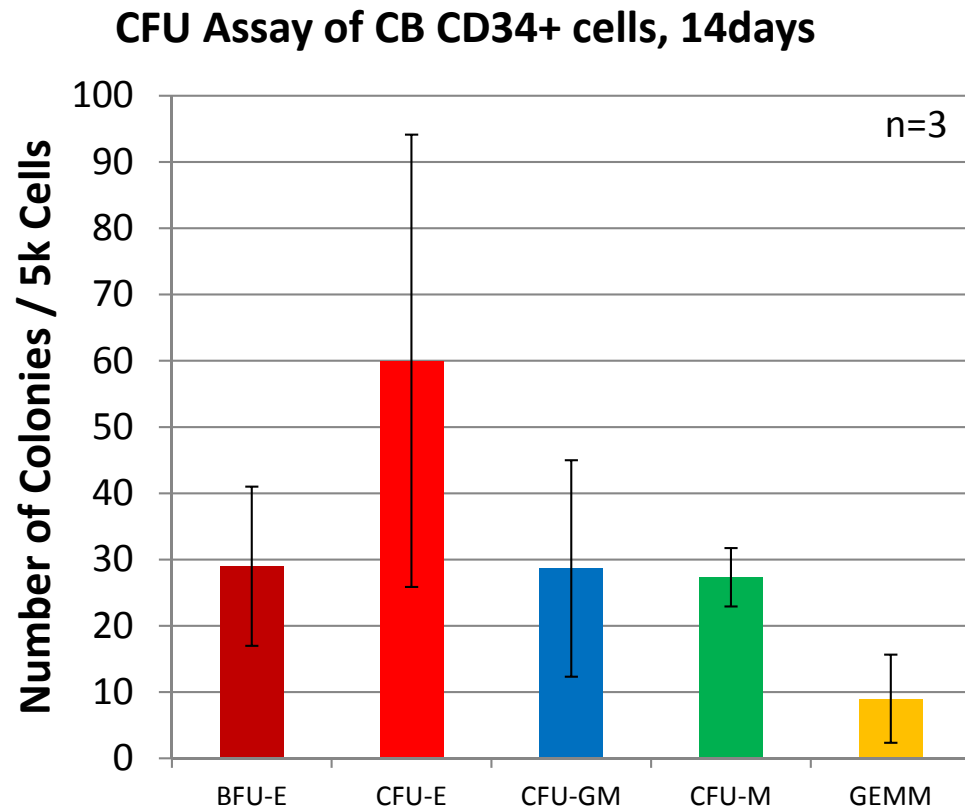


Figure 2-2. Characterization of expanded cord blood CD34+ cells.

CFU assay of expanded CD34+ cells. Data were plotted as mean +/- SEM (n=3).

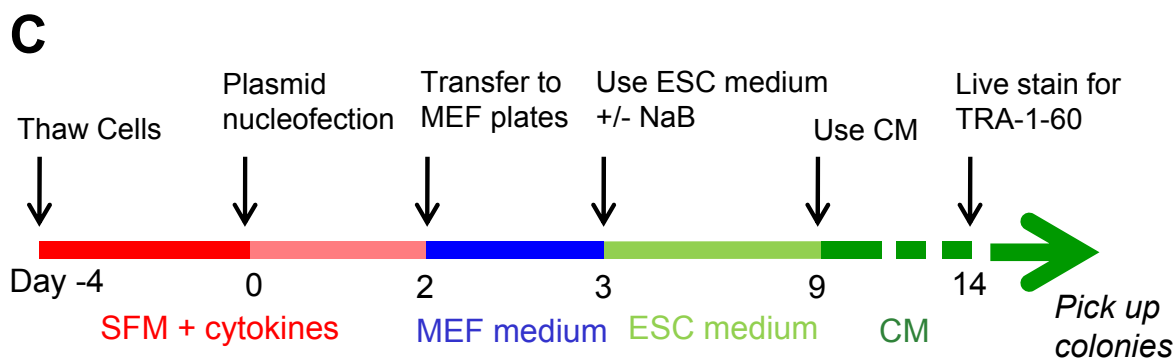


Figure 2-3. Diagrams of novel episomal vectors used in this study, based on an EBNA1/OriP-containing plasmid. (A) A reprogramming vector expressing Oct4, Sox2, Klf4, c-Myc and Lin28 (OSKML) under the control of a chimeric promoter CAG. The plasmid pEB-C5 is also abbreviated later as C5. **(B)** Two supplementary episomal vectors expressing SV40 viral large T antigen (Tg) or an shRNA knocking down p53. In some experiments, pEB-GFP (under the control of the CMV promoter) was used as a vector control, see details in Methods. **(C)** A procedure for reprogramming and identifying iPSC-like cells. Purified CD34⁺ cells were expanded and primed for 4 days, and then nucleofected once by 1 or up to 3 plasmids (a total of 10 μ g DNA) at day 0. Transfected cells were cultured for 2 more days before transferring onto 6-wells plated with mouse embryonic fibroblast (MEF) feeder cells. At day 3, the ESC culture medium was used in the presence or absence of sodium butyrate (NaB, 0.25 mM) or valproic acid (VPA, 0.5 mM). After day 9 when the original MEF feeder cells deteriorated, fresh MEF conditioned medium (CM) was used to substitute plain ESC medium. Starting at day 10, live staining for TRA-1-60 cell surface marker, which is expressed in human ESCs and bona fide iPSCs, could be performed every other day. Typically, we performed TRA-1-60 live staining at day 14, numerated colonies of both TRA-1-60 positive and negative colonies, and picked up colonies afterward.

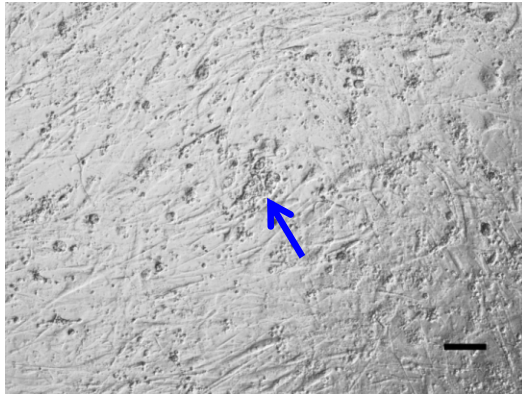
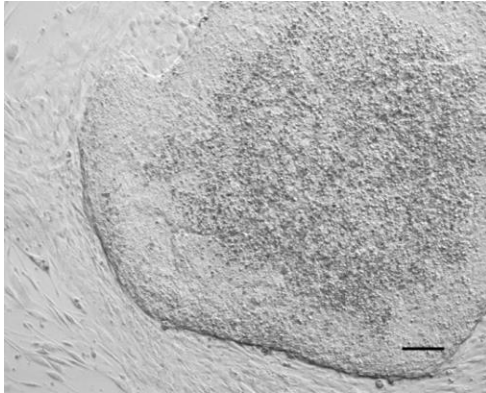
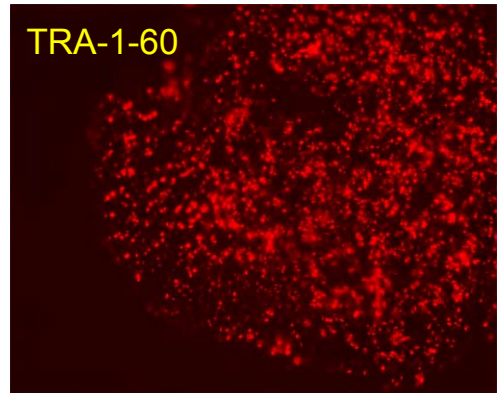
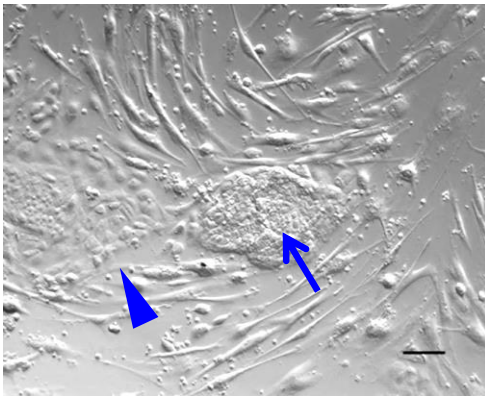
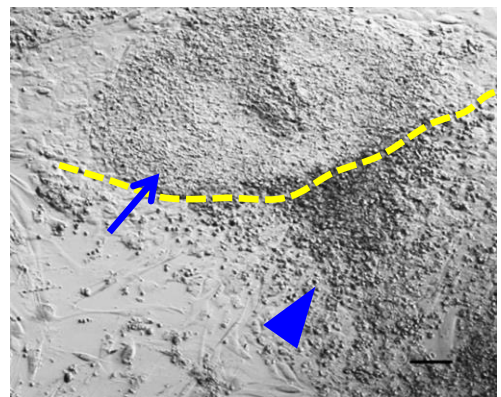
A**B****C****D****E**

Figure 2-4. Morphology of iPSC-like colonies. **(A)** Promising iPSC-like colonies from CB CD34⁺ cell reprogramming can emerge as early as day 6 post episomal plasmid nucleofection, blue arrow. Phase and fluorescence images of a colony after TRA-1-60⁺ live staining 14 days after one-time nucleofection of episomal plasmid **(B)** **(C)** . Nucleofection by the combination #6 which has 3 episomal vectors (Yu et al., 2009) showed partially- and fully- reprogrammed colonies, blue arrow head and arrow respectively on day 10 **(D)** and day 12 **(E)**.

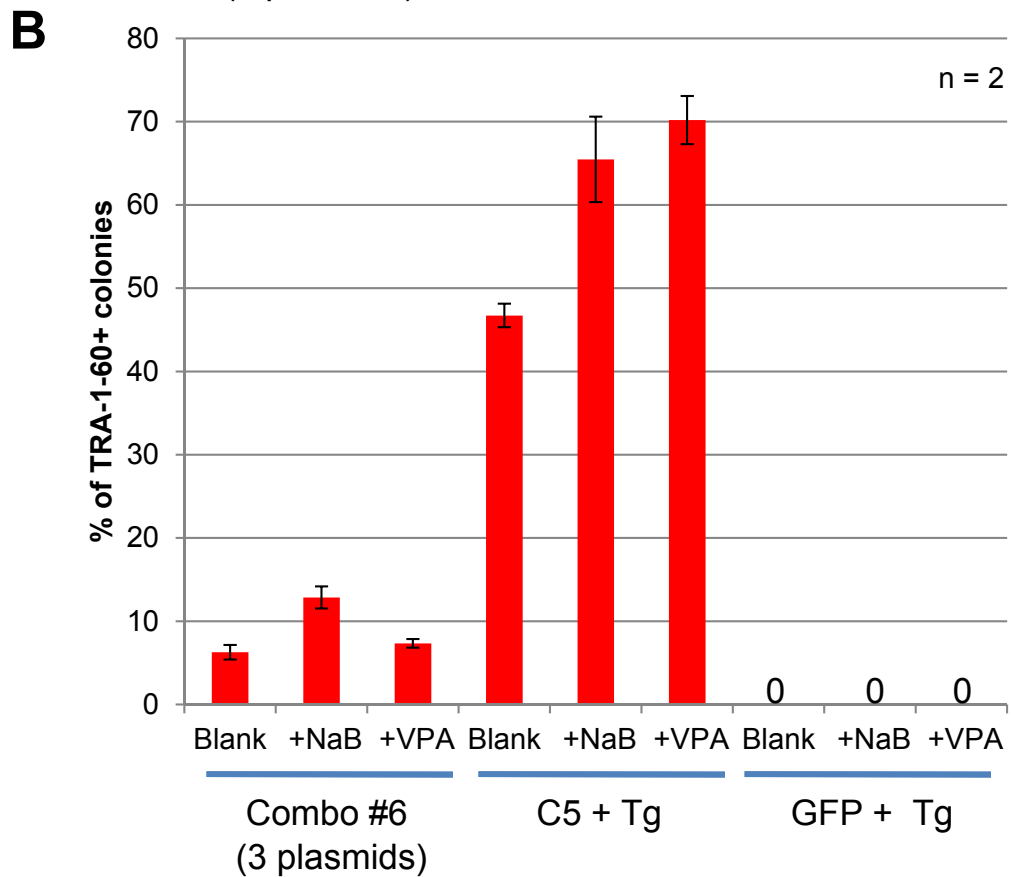
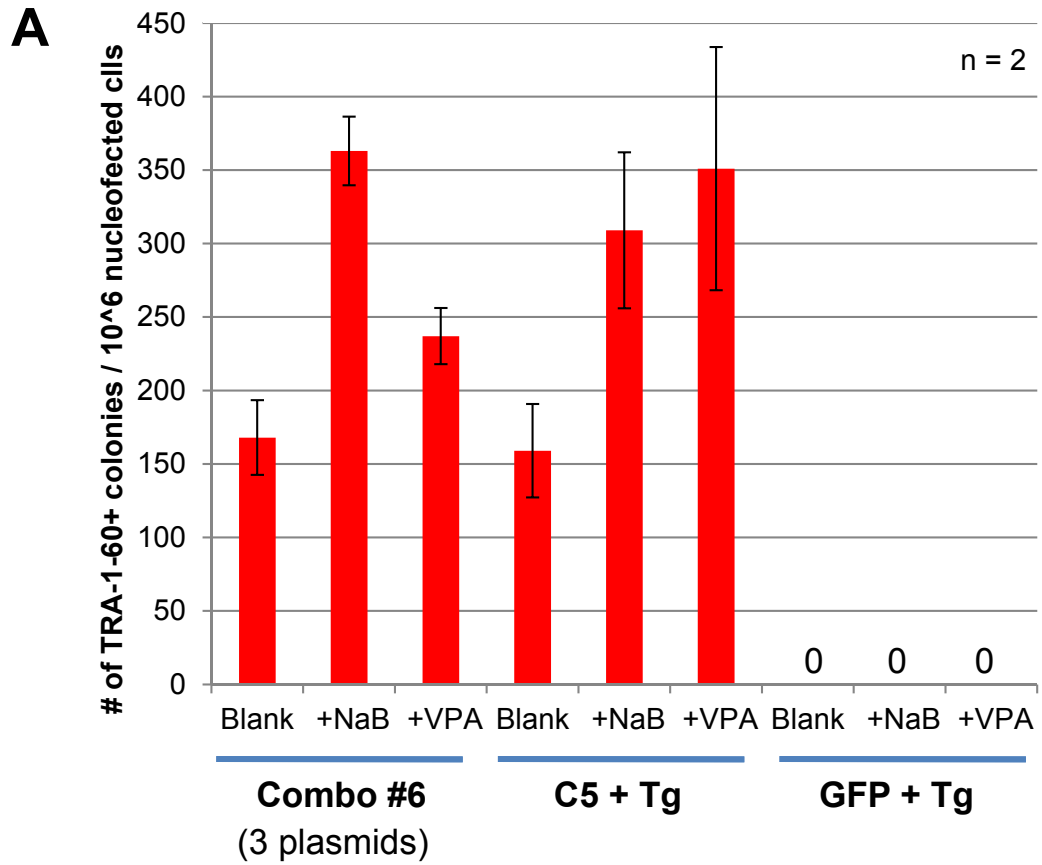


Figure 2-5. Comparison of synergy of pEB-C5 and pEB-Tg episomal plasmids with the combination #6 for cord blood (CB) CD34+ cell reprogramming. (A) Estimating reprogramming efficiency by measuring numbers of TRA-1-60+ and iPSC-like colonies at day 14 per 10^6 nucleofected CD34+ cells. Thomson/Yu combination #6 which contains 3 plasmids expressing 7 genes was compared to the C5+Tg vector combination. Sodium butyrate (NaB) and valproic acid (VPA) were used to test individual additive effect, 0.25mM and 0.5mM respectively. A pEB-GFP (GFP) was used as a negative control for C5. The percentages of TRA-1-60+ colonies among the total emerged colonies were also plotted **(B)**. Data were plotted as mean +/- SEM (n=2).

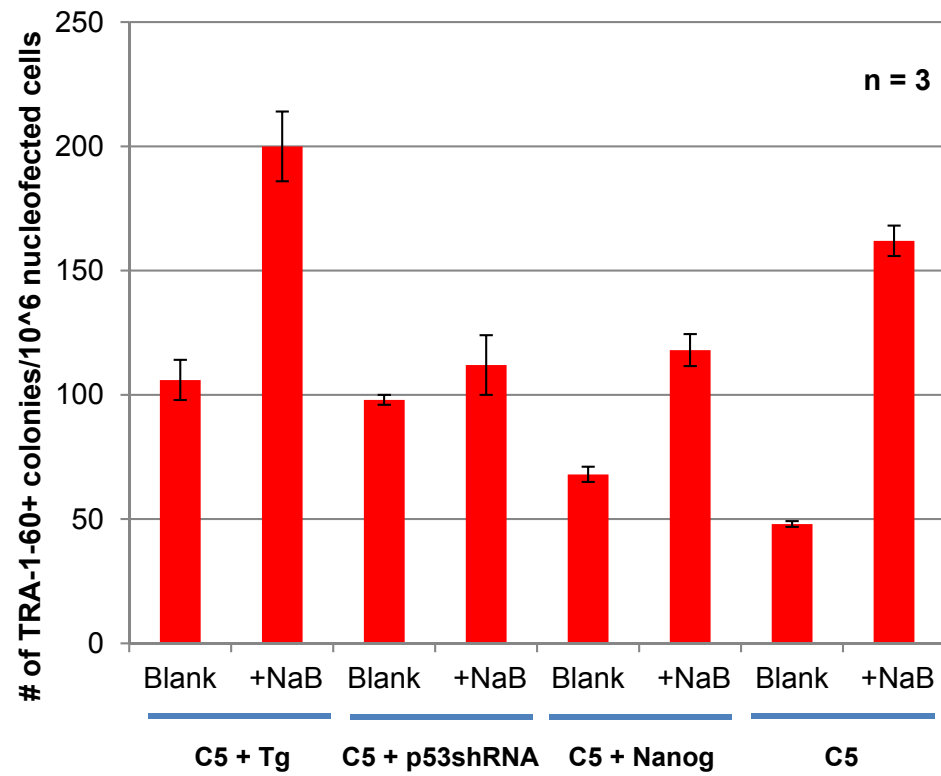
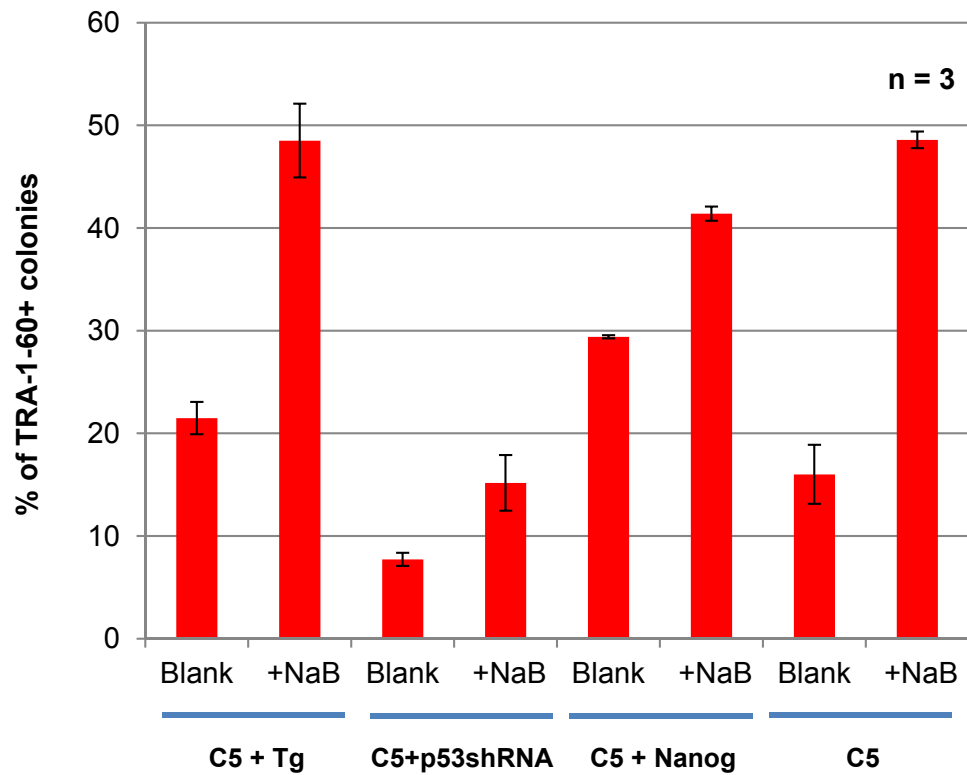
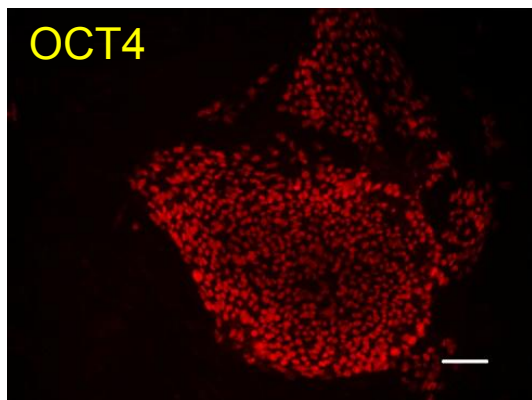
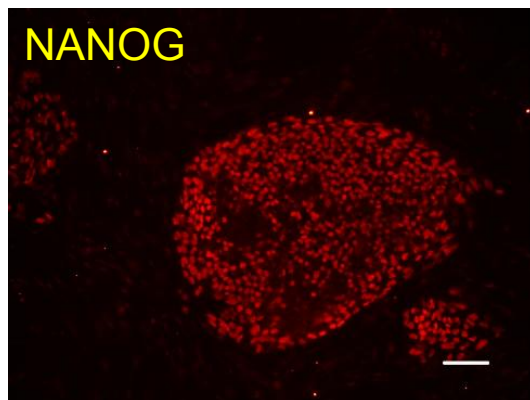
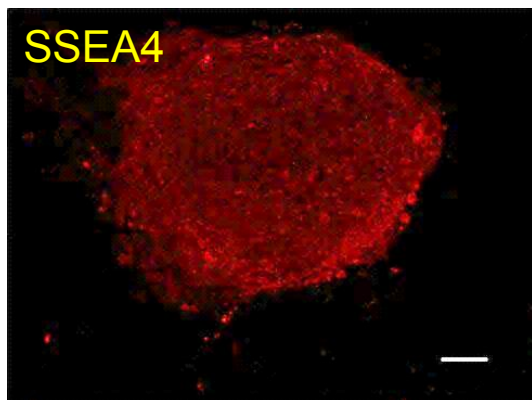
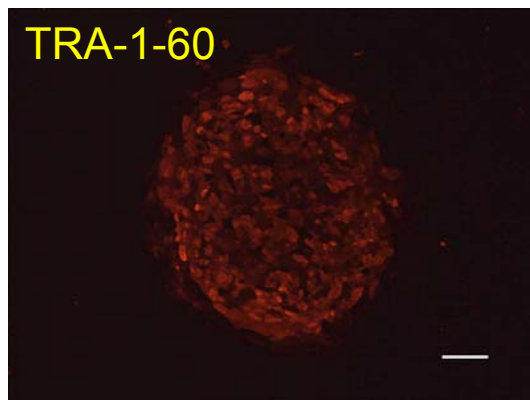
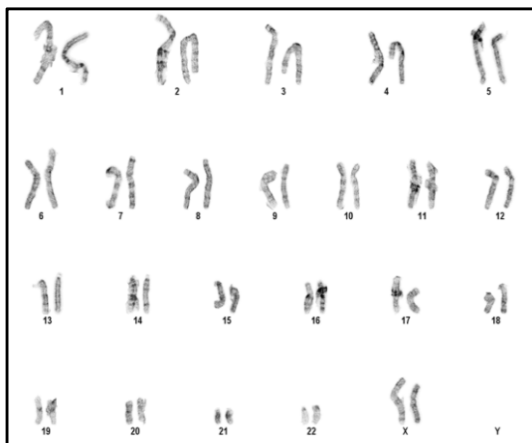
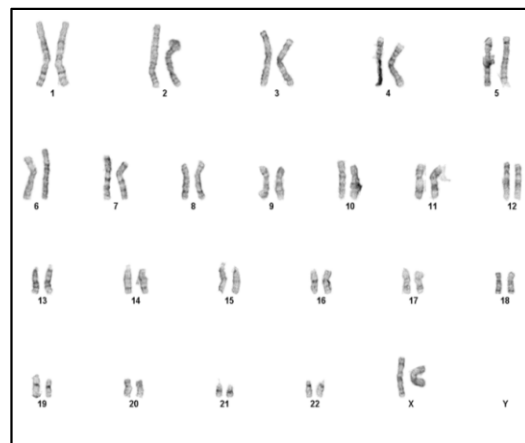
A**B**

Figure 2-6. Reprogramming efficiency of cord blood (CB) CD34+ cells by various combinations of EBNA1/OriP episomal or transient vectors. (A) pEB-C5 (C5) + pEB-Tg (Tg), C5 + pEB-p53shRNA, C5 + NANOG or C5 alone were compared in parallel. The efficiency was measured at day 14 as numbers of TRA-1-60+ and iPSC-like colonies per 10^6 nucleofected cells. Sodium butyrate (NaB) was used to test individual additive effect, 0.25mM. **(B)** The percentage of TRA-1-60+ colonies among the total colonies were also plotted. Data were plotted as mean \pm SEM (n=3). Discrete TRA-1-60+ colonies with the best iPSC-like morphology were selectively picked after counting.

A**OCT4****NANOG****SSEA4****TRA-1-60****B**

C7, p7 (46, XX)



CN1, p7 (46, XX)

Figure 2-7. Characterization of iPSC lines C7 and CN1 derived from CB CD34+ cells by the single pEB-C5 plasmid. (A) Representative images of staining of undifferentiated C7 iPSCs after expansion (more than 5 passages) for markers associated with pluripotency. Upper row for nucleus staining of OCT4 and NANOG. Lower row for surface staining of SSEA4 and TRA-1-60. Immuno-fluorescence staining with specific antibodies followed by Alexa555-conjugated secondary antibodies. Scale bar: 100µm throughout. **(B)** Normal karyotypes of C7 and an additional line CN1 at passages 7, reprogrammed by the pEB-C5 in the presence of sodium butyrate (NaB).

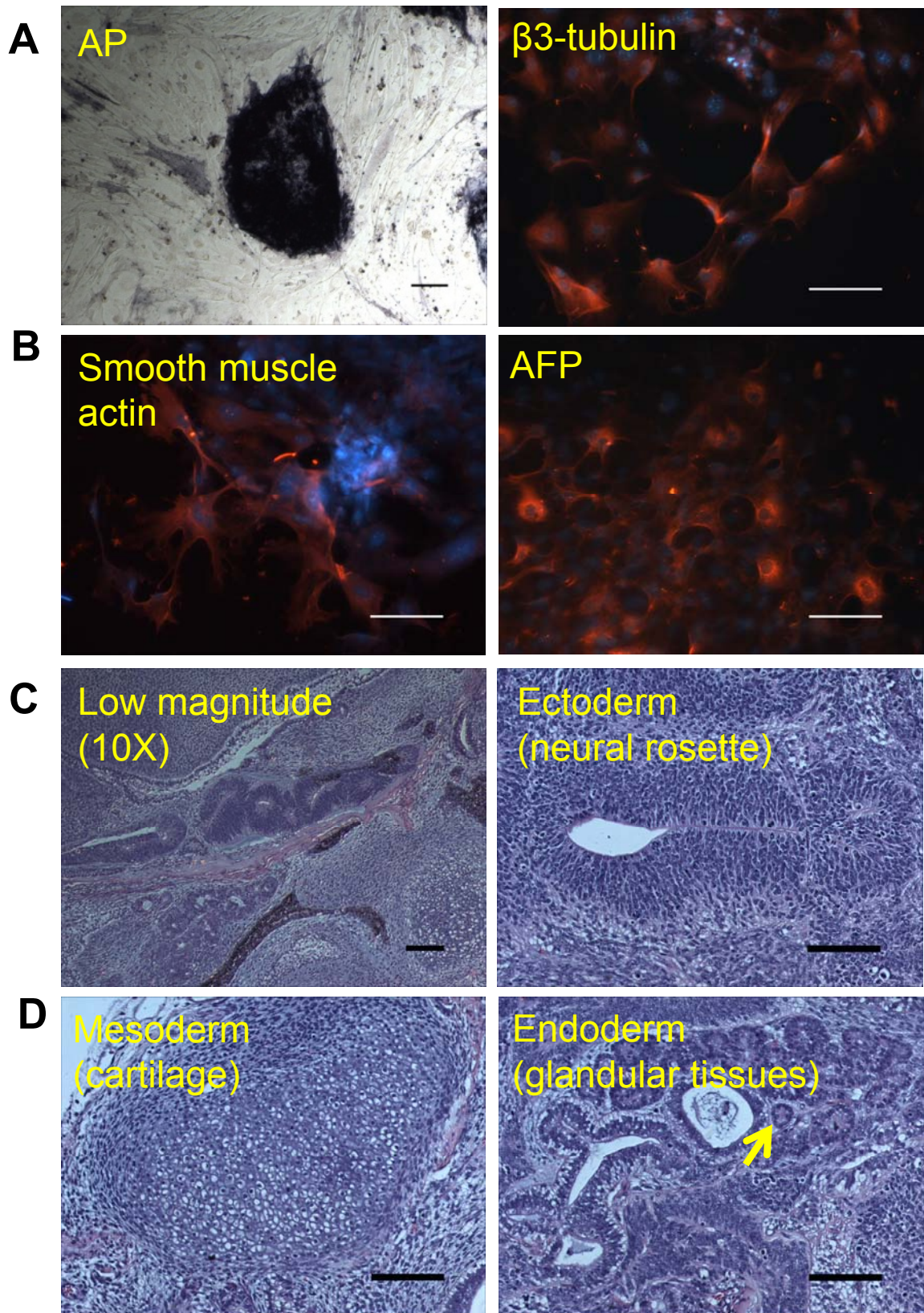
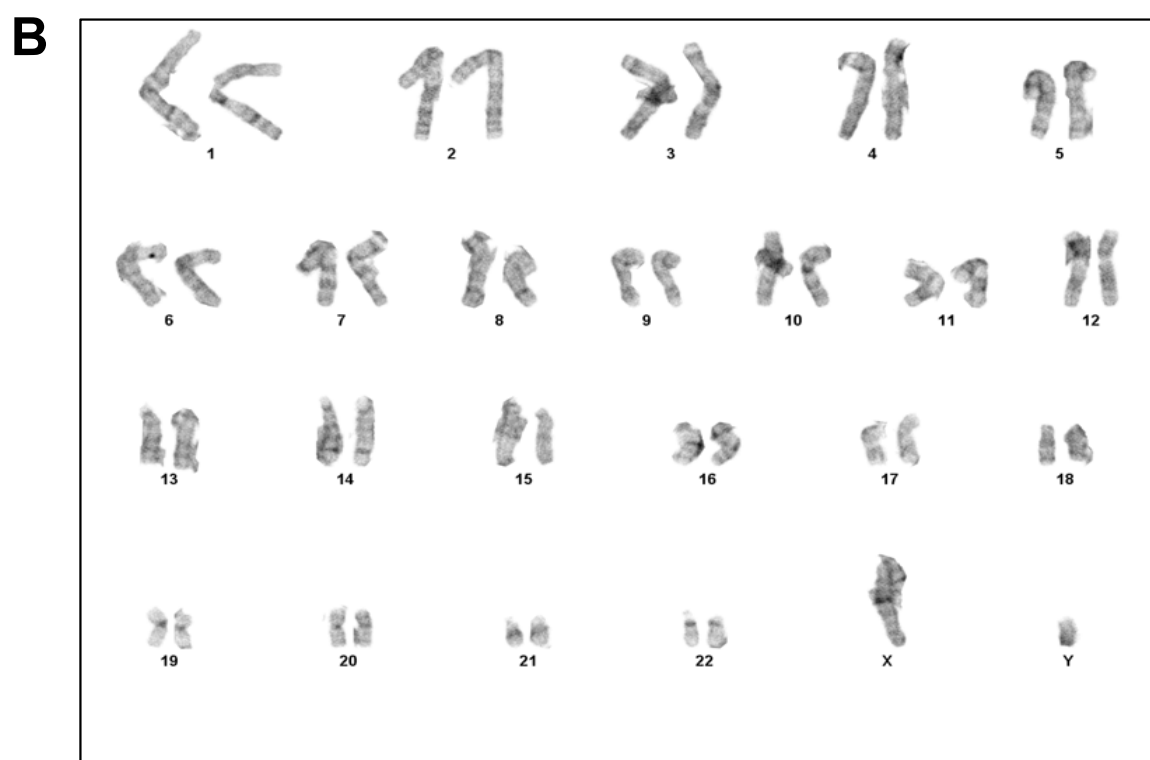
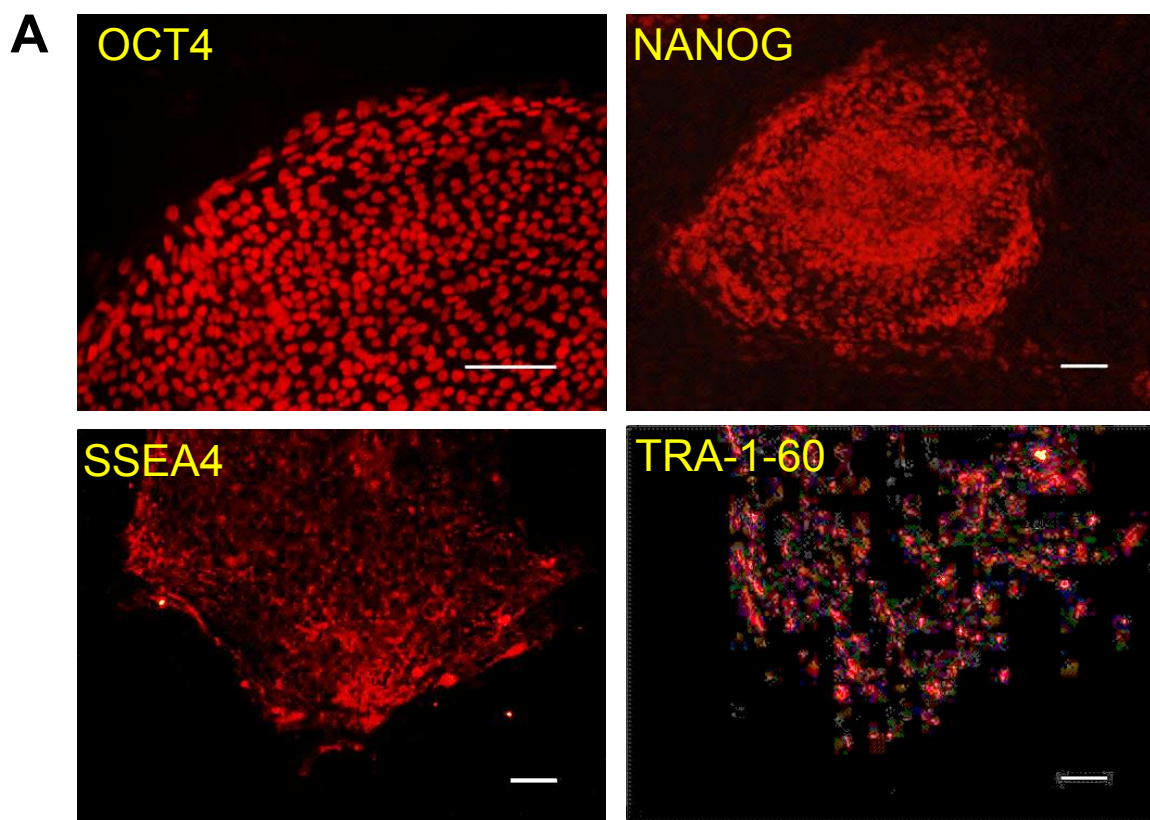


Figure 2-8. Pluripotency staining of iPSC line C7 and *in vitro* and *in vivo* differentiation assay. (A) The C7 iPSCs is positive for alkaline phosphatase (AP) histo-chemical staining, left. (A) right and (B), C7 iPSCs display normal *in vitro* differentiation pluripotency as assayed by embryoid body (EB) formation. Eight days after culture in suspension, cystic EBs were formed and then allowed to adhere for further differentiation. Then the whole culture was stained by monoclonal antibodies recognizing various germ layers including those from ectoderm (β 3-tubulin), mesoderm (smooth muscle actin) and endoderm (alpha-fetal protein or AFP). (C) (D) *In vivo* differentiation assay of C7 iPSCs by teratoma formation. The cystic teratoma-like tumors were excised from animals and sectioned. H & E stained sections were examined. Both low (upper left, 10X) and high (remaining, 20x) magnification images of multiple sections were shown. Various cell types such as neural rosettes (ectoderm) (C right), cartilage (mesoderm) (D left) and glandular structures (endoderm, yellow arrow) (D right) were found. Scale bar: 100 μ m.

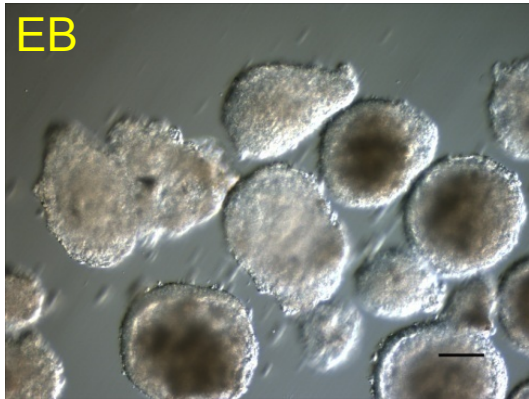


CTN4, p8 (46, XY)

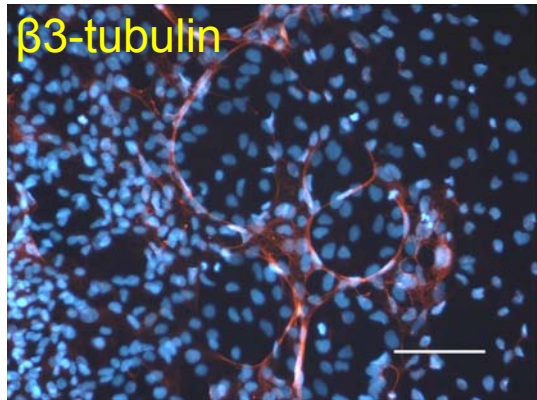
Figure 2-9. Characterization of an additional iPSC line CTN4 derived from CB CD34+ cells by pEB-C5 and pEB-Tg plasmids in the presence of sodium butyrate (NaB). (A) Staining of undifferentiated CTN4 iPSCs after expansion (more than 5 passages) for markers associated with pluripotency. Upper row for nucleus staining of OCT4 and NANOG. Lower row for surface staining of SSEA4 and TRA-1-60. Immuno-fluorescence staining with specific antibodies followed by Alexa555-conjugated secondary antibodies. Scale bar: 100µm throughout. **(B)** Karyotyping of CTN4 at passages 8 revealed normal male.

A

EB

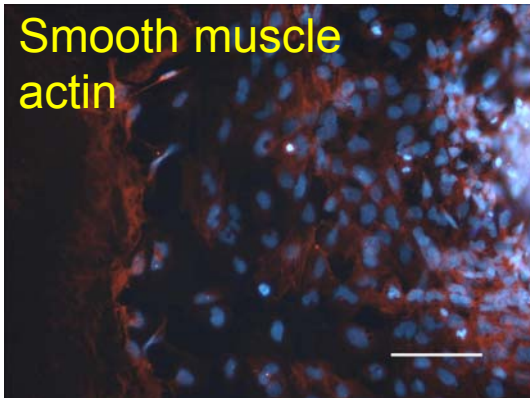


β 3-tubulin

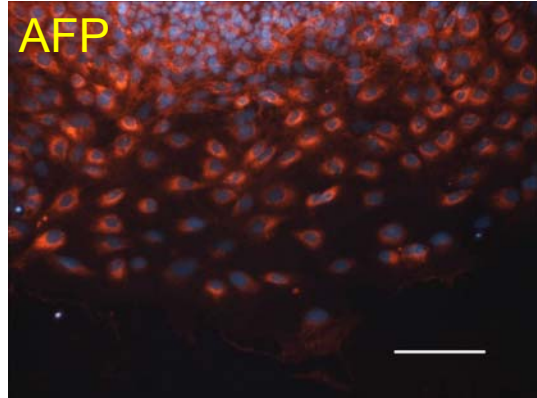


B

Smooth muscle
actin



AFP

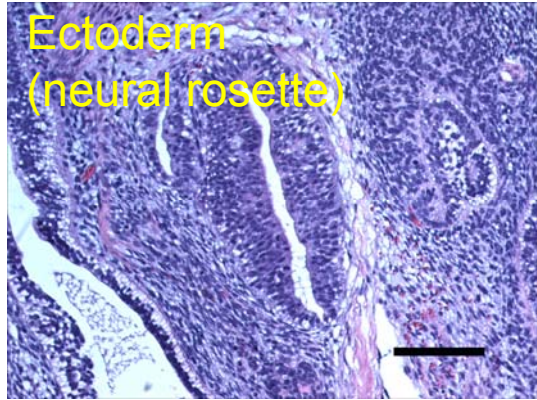


C

Low magnitude
(10X)

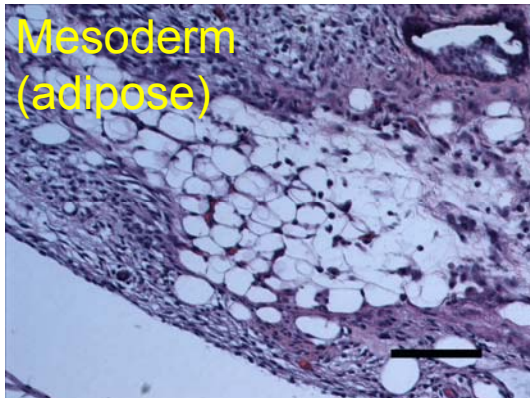


Ectoderm
(neural rosette)



D

Mesoderm
(adipose)



Endoderm
(glandular tissues)



Figure 2-10. Pluripotency assay of iPSC line CTN4 derived from CB CD34+ cells reprogrammed by pEB-C5 and pEB-Tg plasmids in the presence of sodium butyrate (NaB). (A) (B) *In vitro* differentiation assay by embryoid body (EB) formation and differentiation. Various cell types derived from ectoderm (β 3-tubulin), mesoderm (smooth muscle actin) and endoderm (alpha-fetal protein or AFP) were found. Scale bar: 100 μ m. (C) (D) *In vivo* differentiation of CTN4 assayed by teratoma formation. Low magnitude of teratoma were shown in (C left). The presence of various cell types derived from all the 3 embryonic germ layers were shown: neural rosette (ectoderm) (C right), adipose (mesoderm) (D left), and glandular tissues (endoderm, yellow arrow) (D right), in representative images of H & E stained sections. (C left), low magnitude. Scale bar: 100 μ m.

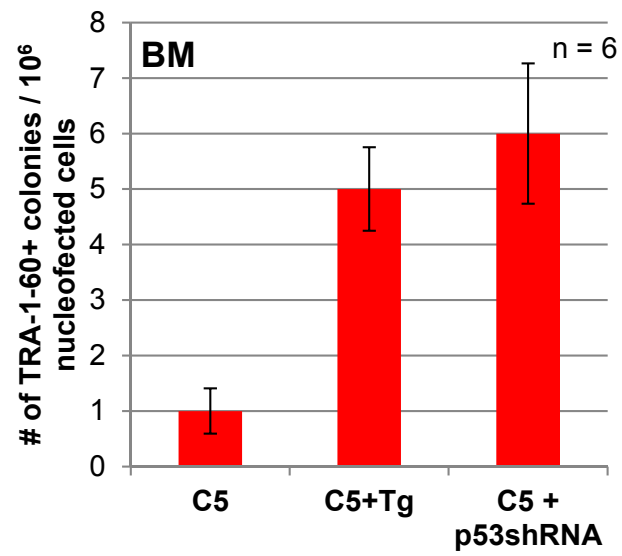
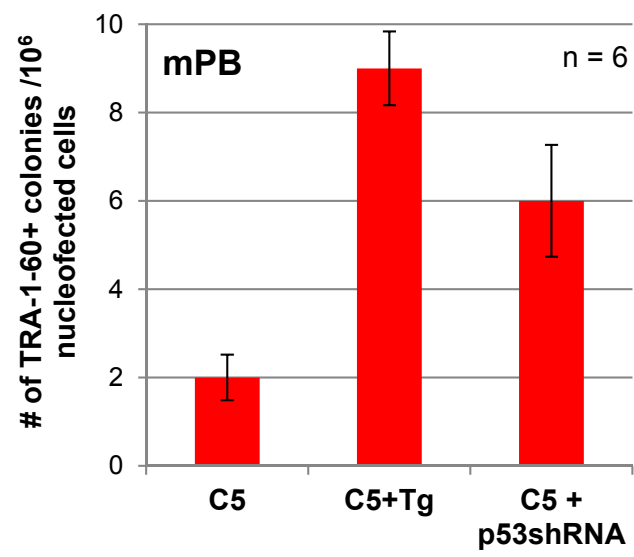
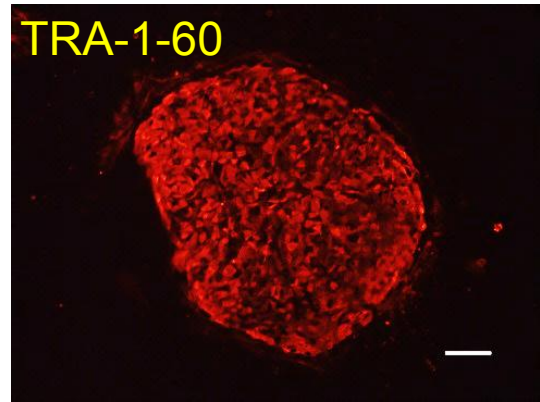
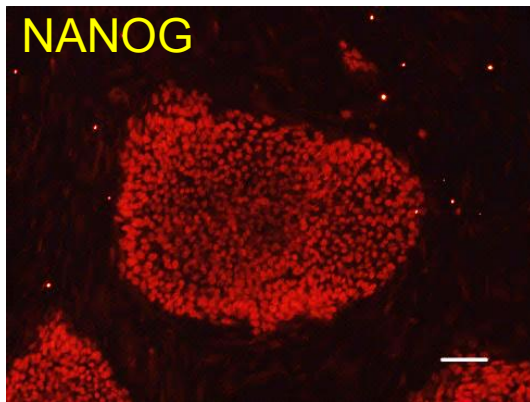
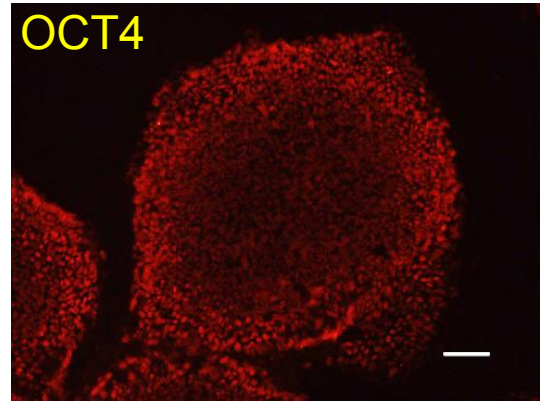
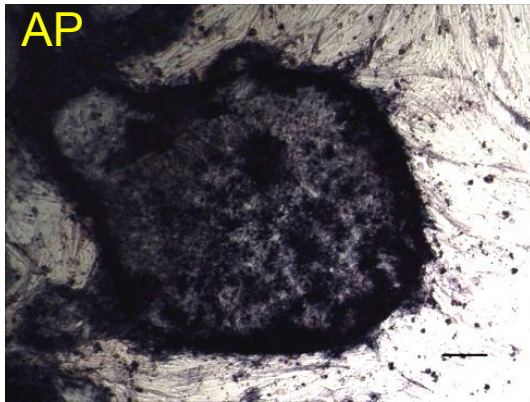
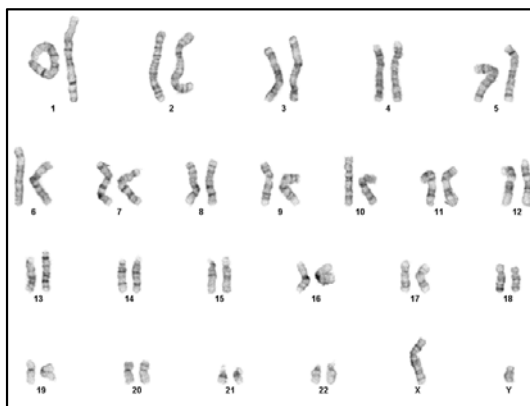
A**B**

Figure 2-11. Reprogramming of human adult hematopoietic CD34+ cells. Reprogramming efficiency of bone marrow (BM, **A**) or of adult mobilized peripheral blood (mPB, **B**) CD34+ cells by various plasmid combinations in the presence of NaB, as conducted before. Numbers of TRA-1-60+, iPSC-like colonies were numerated (mean \pm SEM) at day 14 after one-time nucleofection of 10^6 expanded cells.

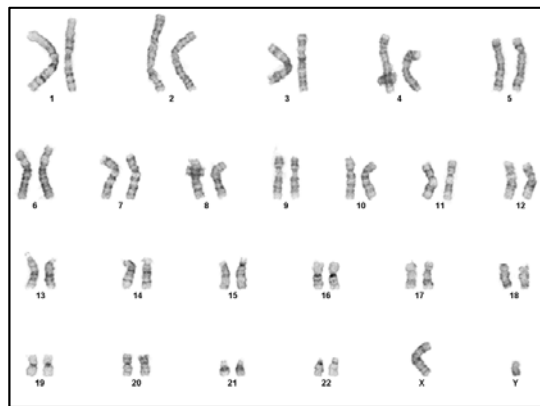
A



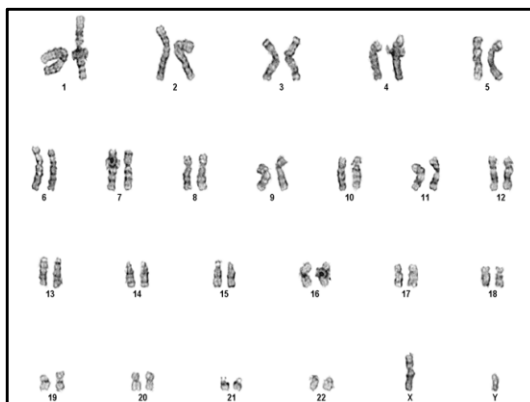
B



BC1, p11 (46, XY)



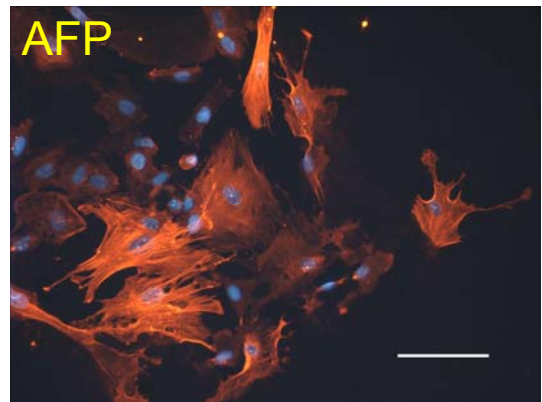
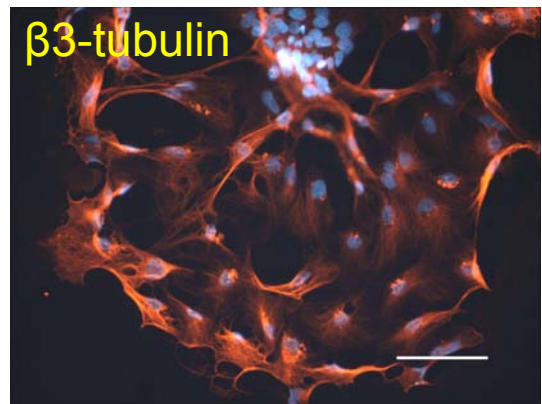
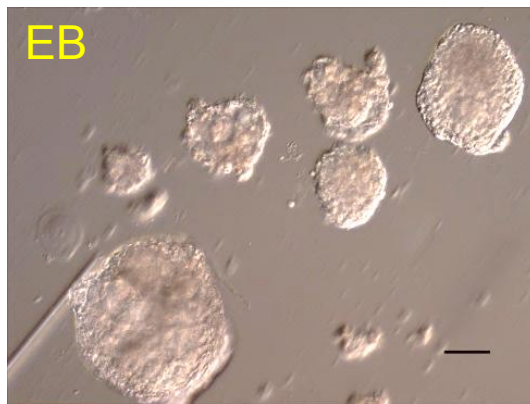
PC1, p12 (46, XY)



BC1, p36 (46, XY)

Figure 2-12. Characterization of iPSC line BC1 derived from bone marrow CD34+ cells by the single pEB-C5 plasmid. (A) Phenotypes of expanded iPSC clone BC1 derived from BM CD34+ cells by the single C5 vector. Upper row for alkaline phosphatase (AP) staining and nucleus staining of OCT4. Lower row for surface staining of SSEA4 and TRA-1-60. Scale bar: 100 μ m throughout. **(B)** Normal karyotypes of iPSC line BC1 at various passages and another iPSC line derived from human peripheral blood CD34+ cells (PC1) by the single pEB-C5 plasmid.

A



B

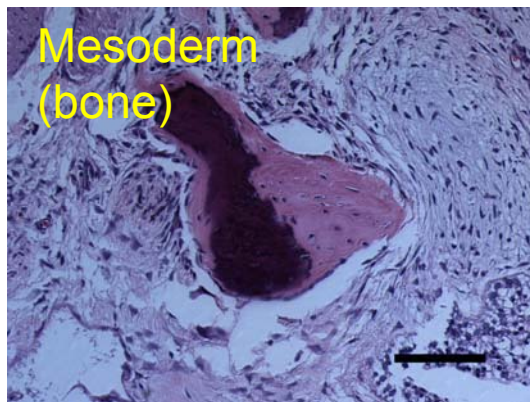
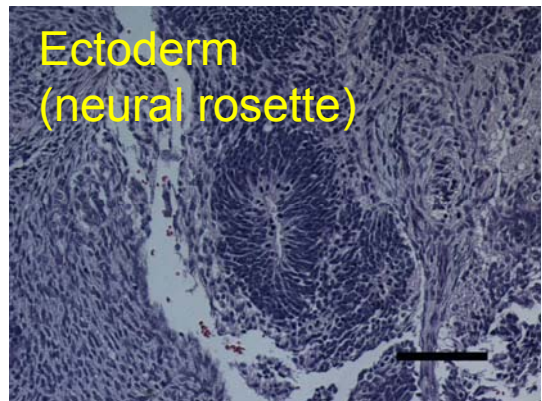
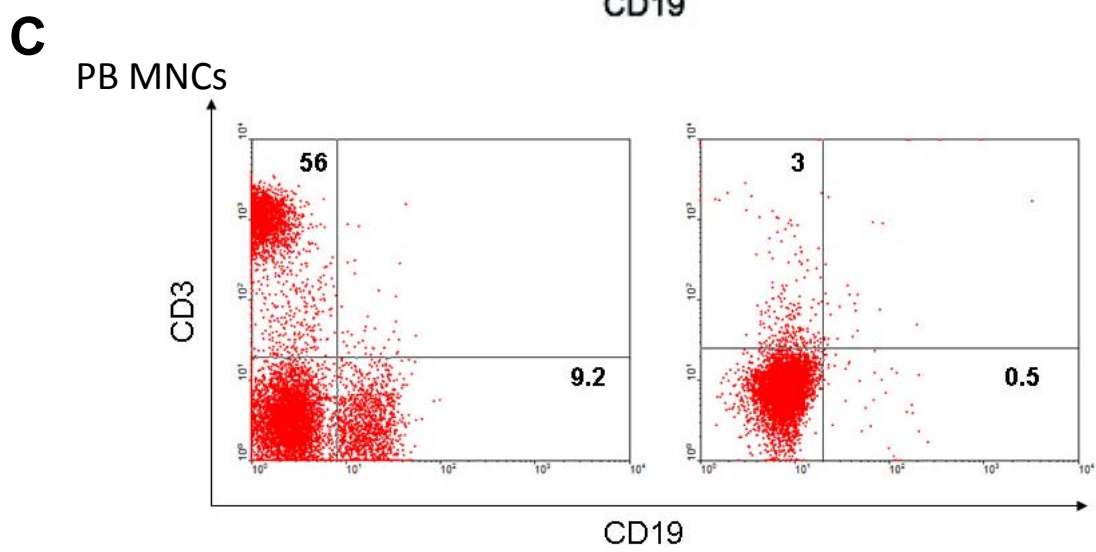
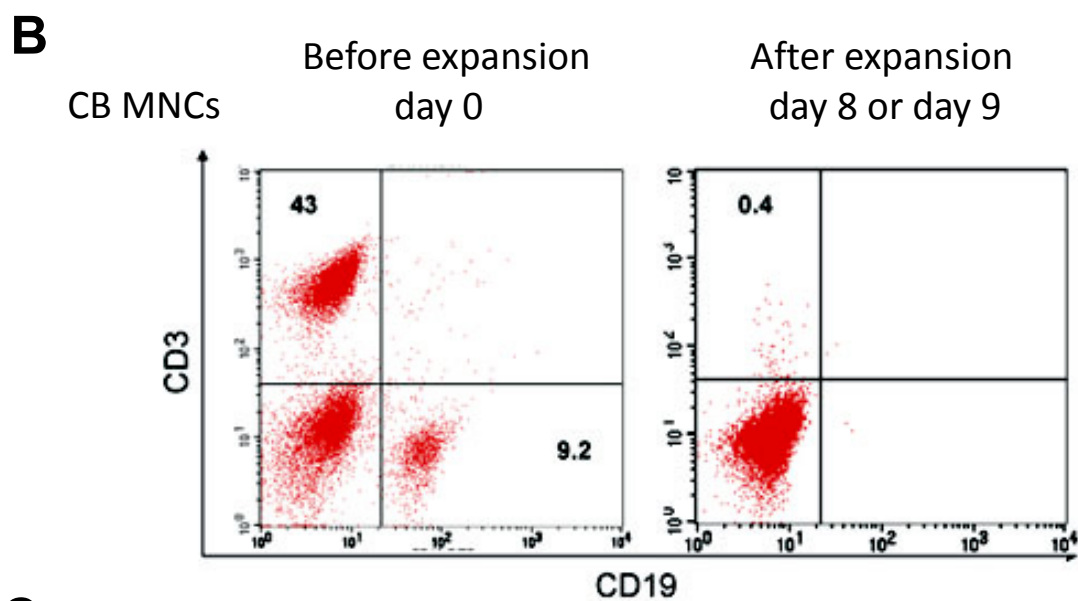
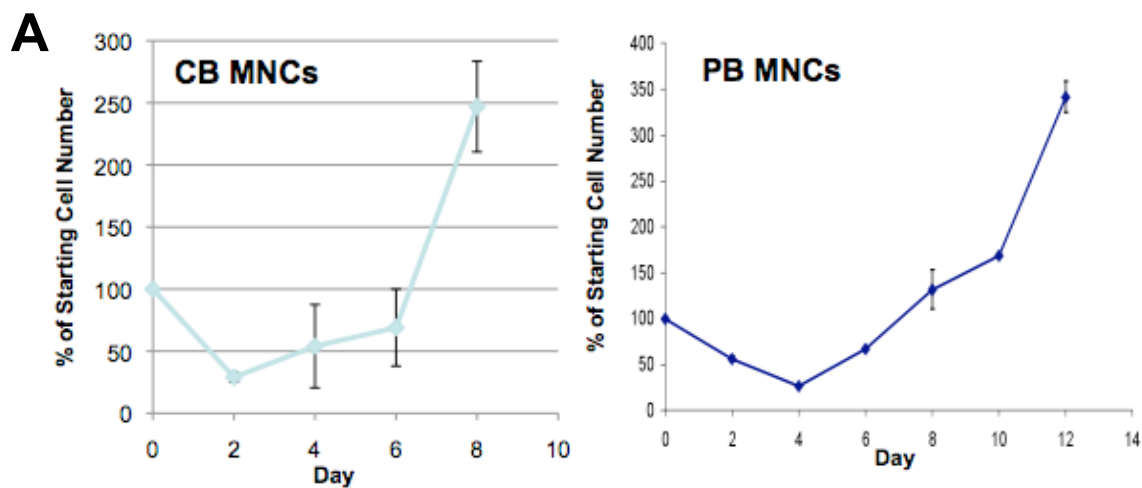


Figure 2-13. Pluripotency assay of iPSC line BC1 derived from bone marrow CD34+ cells by the single pEB-C5 plasmid. (A) The BC1 iPSCs display normal pluripotency as assayed by embryoid body (EB) formation and *in vitro* differentiation. Various cell types derived from ectoderm (β 3-tubulin), mesoderm (smooth muscle actin) and endoderm (alpha-fetal protein, AFP) were found. **(B)** *In vivo* pluripotency of BC1 assayed by teratoma formation. Both low (10 \times) and high (20 \times) magnification images of multiple H & E stained sections were shown. Various cell types such as neural rosettes (ectoderm), bone (mesoderm) and glandular structures (endoderm) were found.



(By Huang, X)

Figure 2-14. Growth curves and phenotypes of expanded mononuclear cells (MNCs) from cord blood (CB) or adult peripheral blood (PB). (A) Plots of viable cells numbers at various days after thaw of CB and PB MNCs at day 0 (defined as 100%). Mean +/- SEM of two samples of CB and PB MNCs were shown. **(B)** Phenotypes of CB MNCs before and after 8 days of culture. **(C)** Phenotypes of adult PB MNCs (from SCD003) before and after 9 days of culture when cells were harvested for flow cytometric analysis of surface marker expression and the remaining cells for reprogramming by nucleofection with 1-2 plasmids. The FACS gating was based on appropriate isotype controls of each monoclonal antibody conjugated with a fluorochrome. After culture (day 8 or 9), very few cells expressed markers for T cells (CD3, <3%), B cells (CD19, <0.5%).

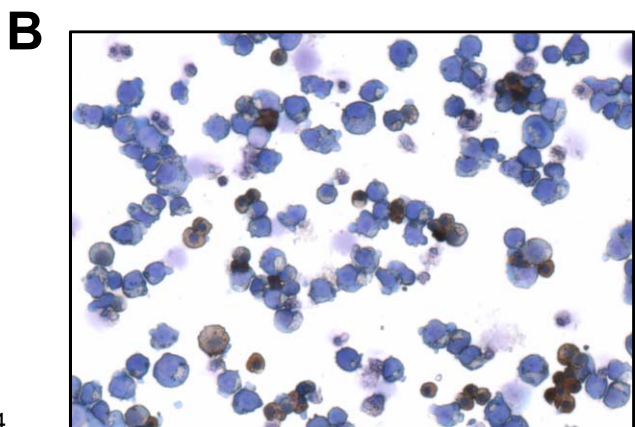
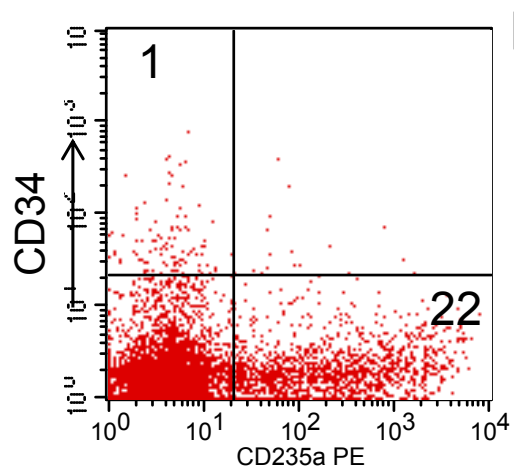
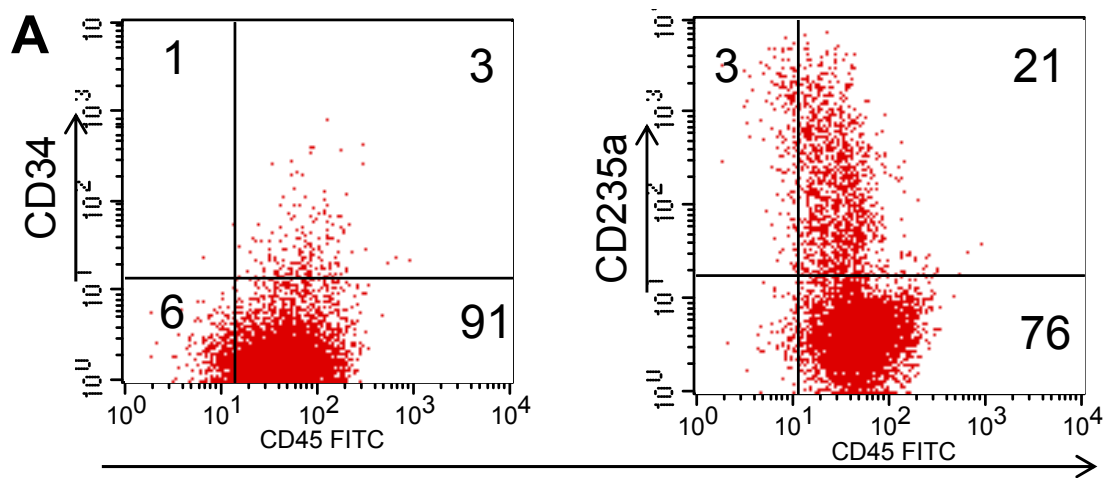


Figure 2-15. Phenotypes of expanded mononuclear cells (MNCs) from newborn cord blood. (A) After 11 days of culture when cells were harvested for flow cytometric analysis of surface-marker expression and the remaining cells for reprogramming by nucleofection with 1-2 plasmids. The FACS gating was based on appropriate isotype-controls of each monoclonal antibody conjugated with a fluorochrome. Percentages of each population in two-color dot plots are indicated. The majority of cells expressed pan hematopoietic cell marker CD45. Some cells resembled erythroblasts that express a high level of CD235a (glycophorin A). Very few cells expressed hematopoietic stem cell marker CD34. **(B)** Morphology of expanded cells from CB MNCs. Benzidine staining of the cells cultured for 16 days, after being spun on slides. Most of cells after the 16-day culture resembled erythroblasts in clusters.

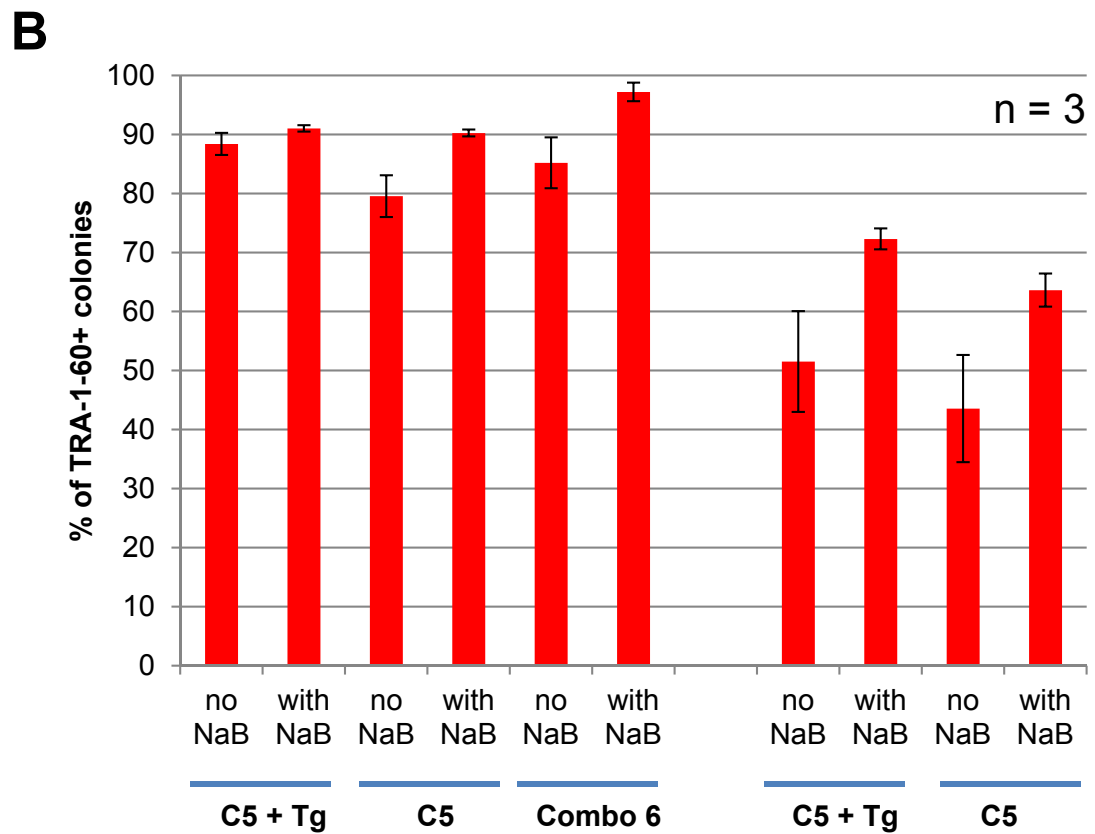
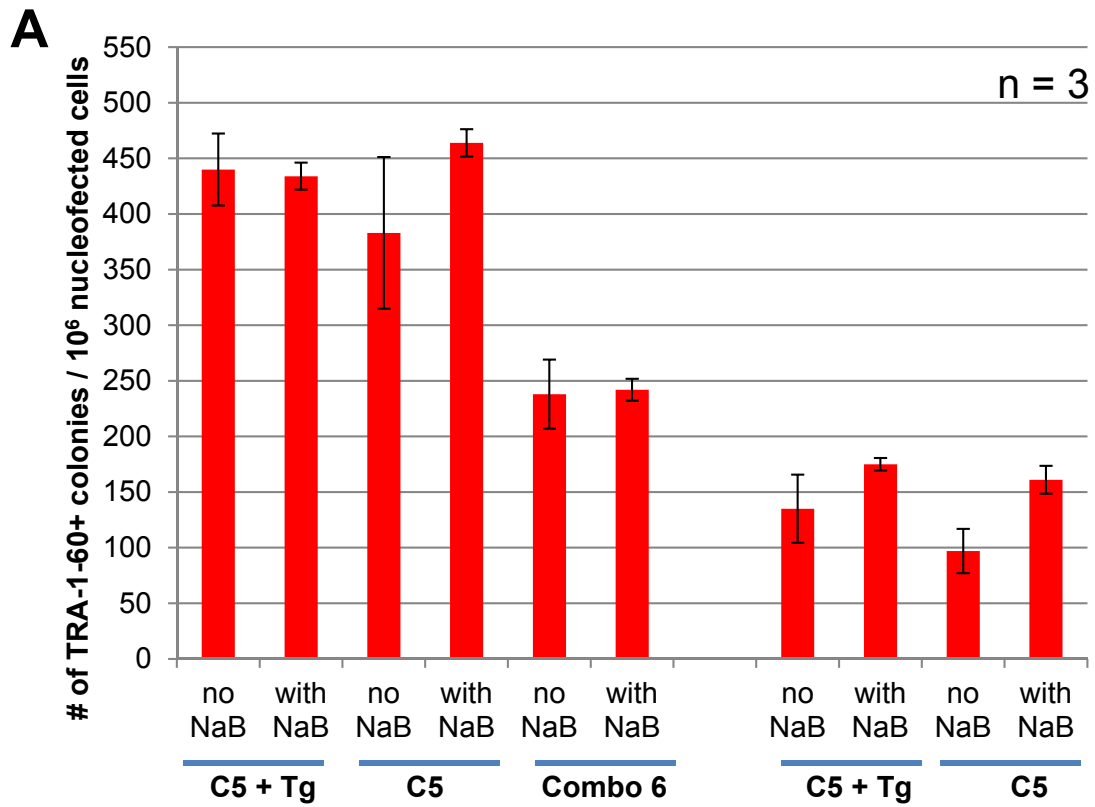
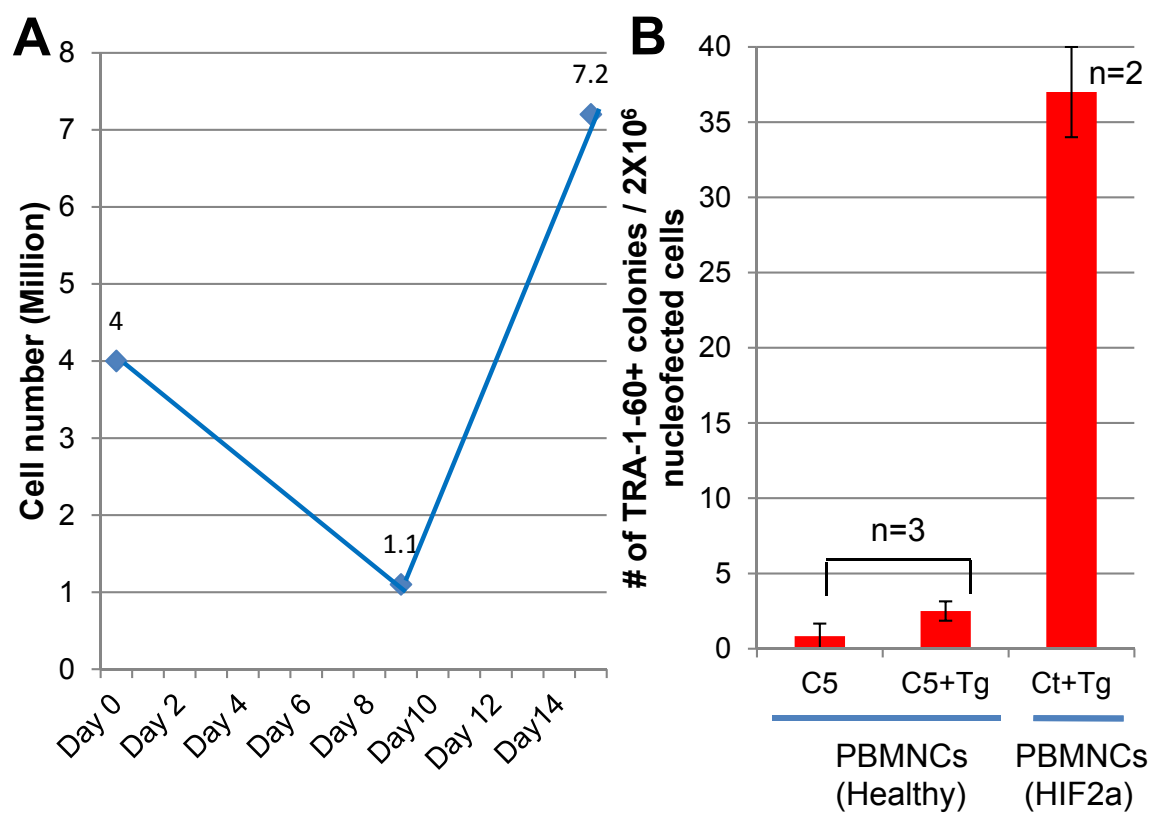


Figure 2-16. Human iPSCs derived from cord blood (CB) mononuclear cells (MNCs) by various EBNA1/OriP plasmids. (A) Reprogramming efficiency of CB MNCs. Two different CB MNCs, frozen 1.5 months before (GB/CB) and 13 years before (AC/CB), were cultured for 8 days. 2×10^6 cultured and primed cells were nucleofected by 1-3 EBNA1/OriP plasmids as indicated. C5: pEB-C5; Tg: pEB-Tg; Combo 6: the Thomson/Yu combination #6 (3 plasmids). The reprogramming proceeded as CD34+ cells described before, with or without sodium butyrate (NaB). Numbers of TRA-1-60+ colonies (on day 14) per 10^6 nucleofected cells were plotted as mean \pm SEM, $n = 3$. **(B)** The percentages of TRA-1-60+ colonies among the total emerged colonies were also plotted.



C

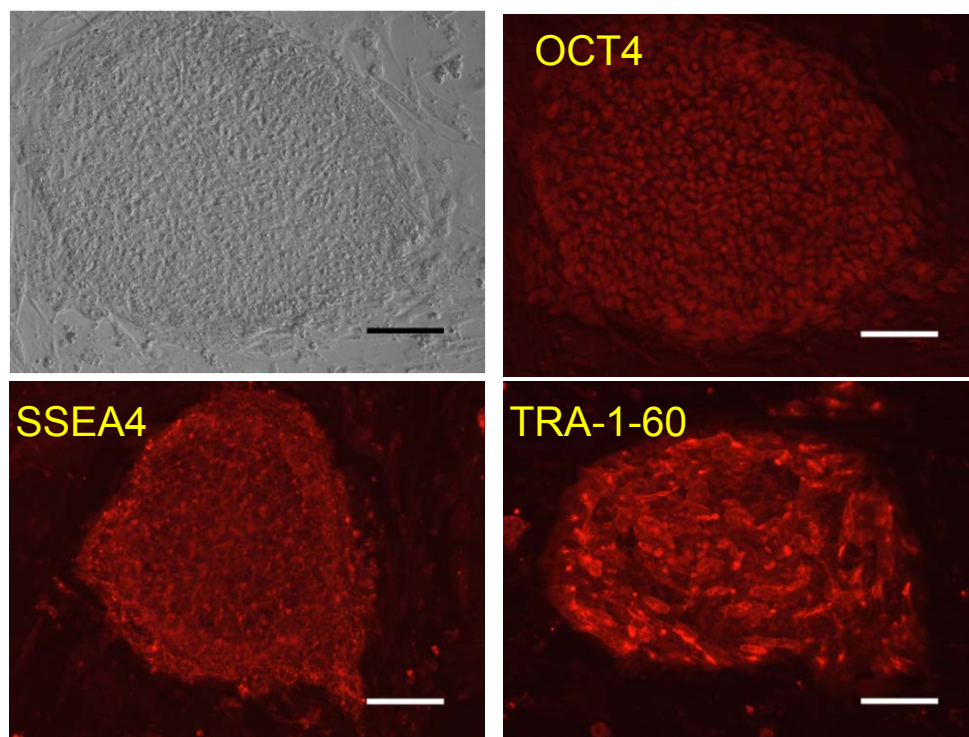
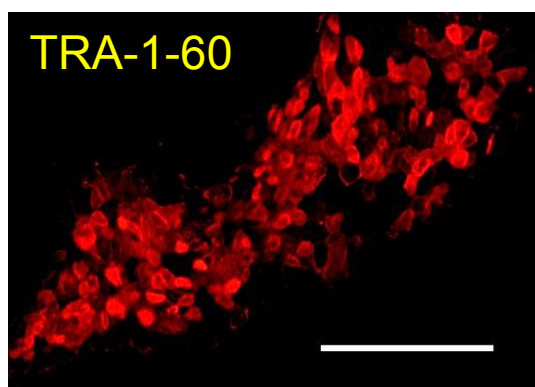
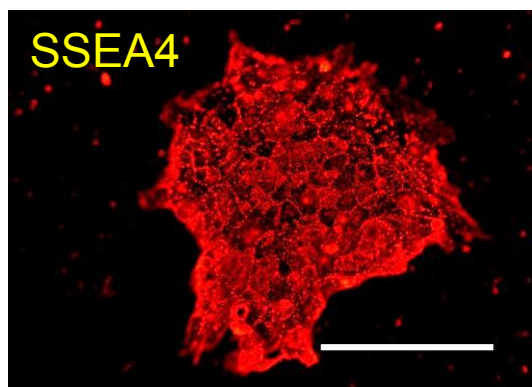
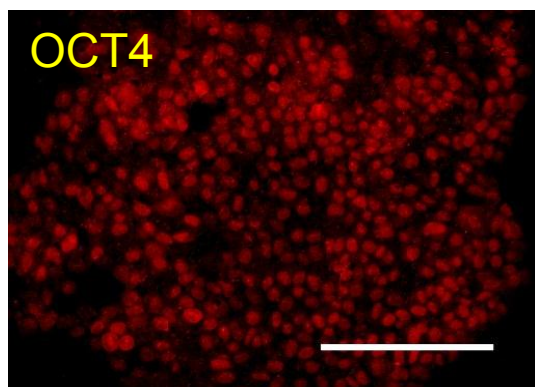


Figure 2-17. Growth curves and phenotypes of expanded mononuclear cells (MNCs) from adult peripheral blood (PB) and resulting iPSC line characterization. (A) Plots of viable cell numbers at various days after thaw of PB MNCs. **(B)** Reprogramming efficiency of PB MNCs. Two different PB MNCs, one from healthy donor and the other one from patient were cultured for 10-14 days. 2×10^6 cultured and primed cells were nucleofected by C5 or together with Tg. **(C)** Phenotypes of expanded iPSC clone PE2 derived from PB MNCs cells by C5 and Tg vectors. Upper row for nucleus staining of OCT4. Lower row for surface staining of SSEA4 and TRA-1-60. Scale bar = 100um.

A



p15 (46, XX)

B

Gly537Arg

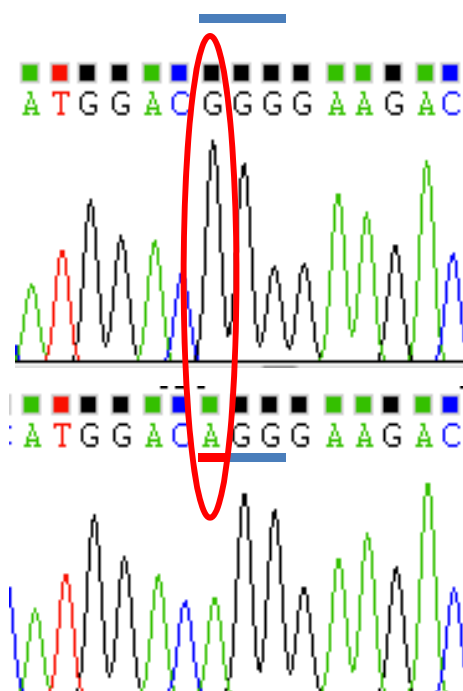


Figure 2-18. Characterization of disease-specific iPSC line from a HIF2 α mutant patient by retroviral vectors. (A) Phenotypes of expanded iPSC clone derived from HIF2 α mutant PB CD34⁺ cells by regular retroviral vectors. Upper row for nucleus staining of OCT4 and surface staining of SSEA4. Lower row for surface staining of TRA-1-60. Normal karyotypes of iPSC line at passage 15. Scale bar: 200 μ m. **(B)** Demonstration of single nucleotide mutation G \rightarrow A in HIF2 α exon 6 by DNA sequencing, causing glycine to arginine change. Scale Bar = 200um.

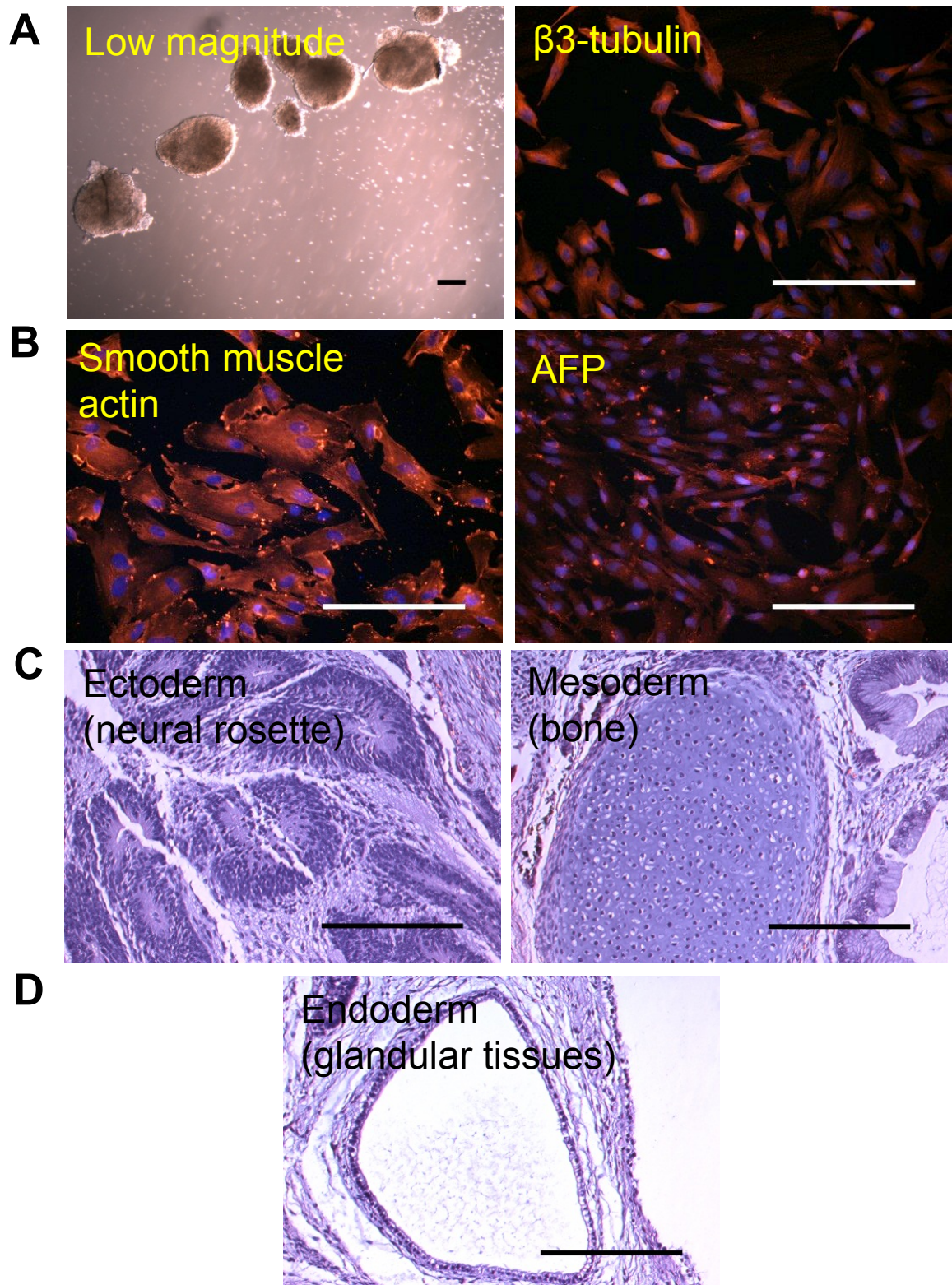


Figure 2-19. Pluripotency assay of iPSC line derived from HIF2 α mutant PB CD34⁺ cells reprogrammed by retroviral vectors in the presence of sodium butyrate (NaB). (A) (B) *in vitro* differentiation assay by embryoid body (EB) formation and differentiation. Various cell types derived from ectoderm (β 3-tubulin), mesoderm (smooth muscle actin) and endoderm (alpha-fetal protein or AFP) were found. Scale bar: 200 μ m. (C) (D) *In vivo* differentiation assay by teratoma formation. The presence of various cell types derived from all the 3 embryonic germ layers were shown: neural rosette (ectoderm), adipose (mesoderm) and glandular tissues (endoderm, yellow arrow) in representative images of H & E stained sections. Scale bar: 200 μ m.

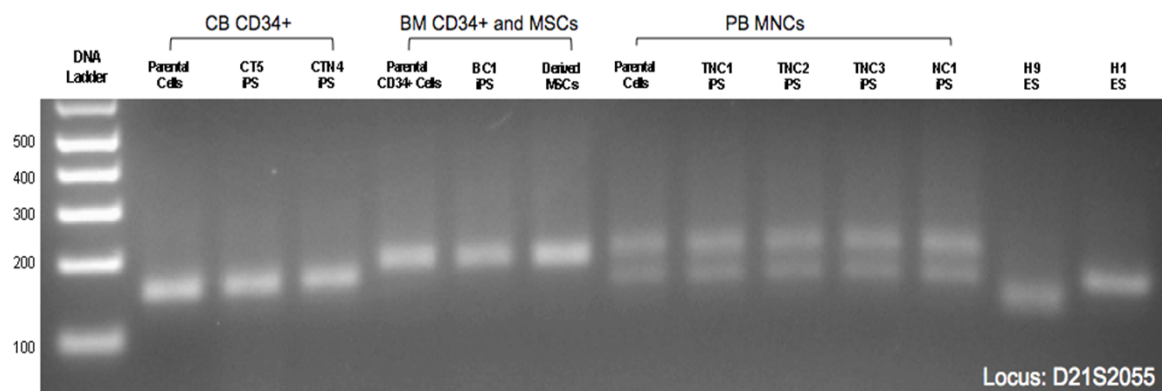
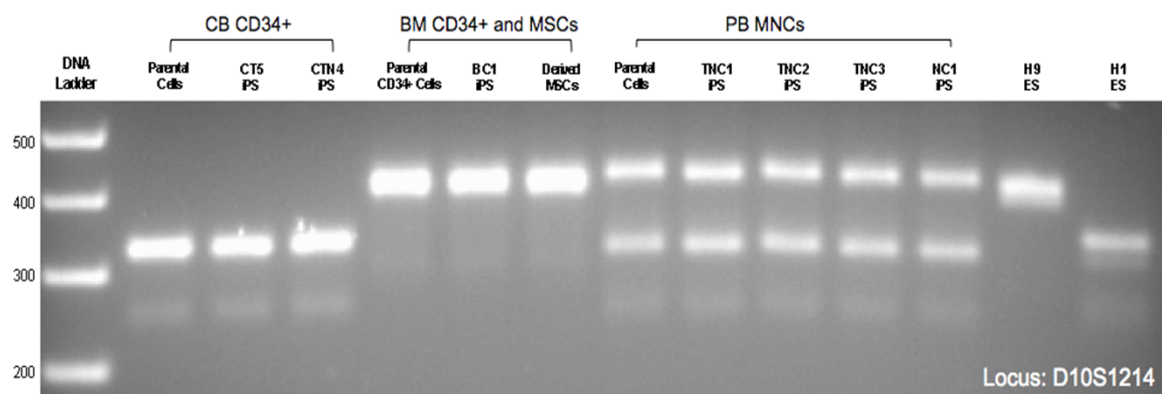


Figure 2-20. DNA fingerprint to confirm the authenticity of iPSCs as compared to their somatic cell origins. Genomic DNAs isolated from various iPSC lines and their parental cells such as cord blood (CB), bone marrow (BM) CD34⁺ cells and MSCs, and adult peripheral blood (PB) mononuclear cells (MNCs). They were used as templates for detection of variable regions (in size) in the genome by PCR using two sets of specific primers from Invitrogen. The PCR fragments were separated by agarose electrophoresis with size markers (in bp). Based on the results using two sets of primers, derived iPSCs showed DNA fingerprints similar to their parental cells, but distinct to other iPSC lines of different origins or two human ESC lines (H1 and H9) we used at the time. For iPSC lines derived from PB MNCs of a sickle cell disease patient (SCD003), we also confirmed that they contain a homozygous mutation at the condon 6 of the β -globin gene (by DNA PCR and sequencing).

Chapter 3

Functional Analysis of the iPSCs Derived from Human Blood Cells by Non-Integrating Plasmids

Background

In chapter 2, I have shown the EBNA1/OriP episomal vector can consistently reprogram various human postnatal blood cells. Compared to standard viral vector-mediated reprogramming, episomal vector has several advantages. First, a single transfection of episomal vector is sufficient for iPSC derivation, no need of multiple treatments or virus packaging. Second, the EBNA1/OriP vectors replicate once in a cell cycle, reducing DNA rearrangement and genome integration (Yates and Guan, 1991). Third, the episomal vectors gradually diminish in human cells, probably due to the silencing of the viral promoter deriving EBNA1 expression in iPSCs (Nanbo et al., 2007). Therefore, simple cell culture and passage may be enough to get rid of residual episomal vectors.

Diminishing of episomal vectors in iPSCs requires many assays to prove it. So in this chapter I will verify the presence of episomal plasmid during and after

reprogramming. It is also equally important to test the pluripotency of iPSCs generated from episomal vectors in a more rigorous way. Since iPSCs were derived from blood cells, it is interesting to examine if blood-derived iPSCs have propensity for hematopoietic lineage re-differentiation.

Approach

The content of this chapter can be divided into 2 parts. The first part is to use multiple approaches to detect the presence of episomal vectors in derived iPSCs. I will start from regular PCR method to detect episomal vectors. I will also use southern blot as an additional method to confirm the result. However, these two approaches are non-quantitative assays. So I will take a quantitative PCR (QPCR) approach to investigate the dynamic change of episomal plasmid number in reprogrammed cells. Tracking the episomal plasmid number in transfected cells at different time points should provide better and more objective evidence. The ultimate method to rule out if episomal vector integrates into the human genome is to harvest both genomic and episomal (from mitochondria) DNA from derived iPSCs and undergo whole genome sequencing (WGS). By analyzing WGS, I can gain insight the integrity of the genome before and after reprogramming.

The second part is to characterize the pluripotency of iPSCs derived by non-integrating plasmid. To collaborate with other researchers, I conducted direct differentiation of iPSCs to various lineages and confirmed the differentiation potential of derived iPSC clones.

Results

PCR to Detect the Existence of Episomal Vectors in Derived iPSCs

Although EBNA1/OriP episomal vectors can reprogram a broad spectrum of blood cell types, the assumption of integration-free has not been validated. To this end, I extracted genomic DNA from cells before reprogramming (naïve cells) or cells 2 days post-nucleofection as negative and positive controls. Genomic DNAs from three different iPSC lines were tested: BC1 and PC1 are from BM and PB CD34+ cells respectively and C7 is from CB CD34+ cells (**Figure 3-1**). I designed primers specific to EBNA1 or Tg sequence to detect any residue of vector DNA using PCR. In general, a trace amount of EBNA1 DNA sequence (<0.2 copies per cell) could be detectable in iPSCs (C7) after reprogramming and expansion of 9 passages (about 50 days), but became undetectable by passage 10-12 (all three lines, **Figure 3-1A**). Since all lines were reprogrammed by C5 plasmid alone, no Tg sequence should be detected.

Two characterized iPSC clones, CT5 and CTN4, which reprogrammed by C5+Tg were also analyzed by PCR to detect the presence of vector DNA EBNA1 and Tg in expanded iPSCs at various passages. I again found that the plasmid DNA was undetectable after 10-12 passages of iPSC expansion (**Figure 3-1B**).

I further used a third pair of primers which is specific to the junction of Sox2/Klf4 transgenes (in C5 plasmid) to check BC1 and 2 additional iPSC lines – TNC1 and NC1. TNC1 and NC1 cell lines were reprogrammed from adult PB MNCs. In line with previous observations, I did not detect any sequence from

Sox2/Klf4 junction. My data are consistent with previous publications with reprogrammed human fetal or neonatal cells that the EBNA1/OriP episomal DNA in human iPSCs was spontaneously and gradually lost after reprogramming (Yu et al., 2009; Marchetto et al., 2009), likely due to methylation at EBNA1 and/or OriP DNA sequences resulting in silencing of EBNA1 expression and/or loss of OriP functions.

Southern Blot Demonstrated the Absence of the Episomal Vectors

I also conducted Southern blot analysis that is capable of detecting a single copy of vector DNA (**Figure 3-2A, B**). I found that the southern blot did not detect the vector DNA either as episomes or anywhere in the genome of NC1, TNC1, BC1 and C7 iPSCs. **Table 3-1** summarizes the characterizations of 10 most analyzed iPSC lines I derived from human blood cells by transient transfection of 1-2 episomal vectors.

QPCR to Detect the Dynamic Change of Episomal Plasmid Numbers per iPSC at Different Passages

Standard PCR only provides information of episomal plasmid presence/absence in established and expanded iPSC clones. To quantitatively measure the dynamic change of episomal vector in reprogrammed cells during the reprogramming and iPSC expanding procedure, I took the quantitative PCR (QPCR) approach. Two PCR primer pairs were designed; one is specific for EBNA1 sequence (EBNA1-primer) and the other one can recognize a small

fragment spanning the junction of Oct4 and Sox2 transgenes (OSJ-primer) (**Figure 3-3A**). EBNA1-primer or OSJ-primer was first used to run QPCR with different amounts of pEB-C5 plasmid and plotted the threshold cycle for the standard curve reactions against the fold dilution of the template C5 plasmid (**Figure 3-3B**). I can then determine the fit to a straight line including the slope and the correlation factor (**Figure 3-3B**).

Next I performed another round of CB CD34+ reprogramming using pEB-C5. Genomic DNA was extracted from day 2 nucleofection cells or derived iPSCs of various passages (**Table 3-2**). I loaded 100 ng of gDNA of different samples for QPCR reaction with either EBNA1- or OSJ-primer then got the Ct value and calculated corresponding episomal plasmid number. **Table 3-2** was the QPCR result of plasmid copy numbers per iPSC at different passages. In the day 2 sample, I can detect 363 copies of plasmid per cells. However, by iPSC passage 6 or higher, the plasmid number decreased to 0.002 copies per iPSC. The data suggest that most of episomal plasmids were already diluted out. Higher passages of iPSCs also showed very low copy number of episomal plasmid.

Whole Genome Sequencing Revealed Absence of Episomal Vector Sequence in the Genomes of the iPSC Lines

PCR-associated assays provided a quick examination about existence of vector sequence C5 or Tg. But it is not the ultimate method especially if I want to look into the DNA level before and after reprogramming. With recently available technologies such as whole-genome sequencing (WGS) analysis, it is necessary

to assess potential vector integration in the entire nuclear and mitochondrial genomes. I conducted WGS analysis to determine the DNA sequences of two human iPSC lines.

The relationship of the two iPSC lines and their somatic cell ancestors are shown (**Figure 3-4**). BM CD34⁺ cells were used to generate the BC1 iPSC line after 4 day culture with pEB-C5. The second iPSC line BCT1 was derived from the same cultured CD34⁺ cells by C5 together with Tg. For the whole-genome DNA sequencing, expanded iPSC lines were cultured under a feeder-free condition (on matrigel) for at least one passage to reduce mouse feeder cells before total genomic DNA was extracted. In addition, DNA was extracted by the same method from the corresponding parental cells, i.e., the cultured CD34⁺ cells. My colleagues and I sequenced the iPSC lines BC1 and BCT1 with their parental somatic cells at the same time in a pair-wise fashion by Illumina HiSeq 2000 technology. We obtained alignable DNA sequences around 130 Gbs for each iPSC line and their parental cells. The amount of sequence data is equivalent to 48 times or higher coverage of the haploid, nuclear genome. Sequence analysis indicated that we achieved high-quality coverage for more than 94% of the autosomal genome of each sample. This level of deep and pair-wise sequencing of three related genomes (from the same person) provided us a high level of confidence to detect any integration of vector sequences.

The deep sequencing of the total DNA (5 µg or from 800,000 iPSCs) provided more definitive evidence for the lack of vector DNA (either integrated or as episome) in these iPSC lines. I did not detect the pCEP4 vector backbone

sequence above background in any of the two sequenced iPSC lines derived by episomal vectors. **Table 3-3** shows a summary of sequence analyses of the iPSC lines.

Analysis of DNA Sequence Variations in Derived iPSCs in Comparison with Parental Somatic Cells

When comparing to the human reference genome sequence (hg19), we identified approximately 4.2 million single-nucleotide variants (SNVs) in each of the iPSC lines as well as in their parental CD34+ cells and MSCs (**Table 3-3**). Some of these SNVs were aligned to repetitive regions of the reference genome (in parentheses). This level of sequence variations is comparable to those of other sequenced human genomes (Bentley et al., 2008; Ding et al., 2010; Lee et al., 2010; Pleasance et al., 2010; Ajay et al., 2011). When the sequences between each iPSC line and its parental cells were directly compared, 1,058 to 1,808 likely SNVs were identified between each iPSC line and its parental somatic cells sequenced in pairs (**Table 3-3**). All SNVs found in iPSCs were heterozygous (i.e., single-allele) changes as compared to their parental somatic cells. Importantly, none of these iPSC-associated SNVs are shared among the three iPSC lines. Neither did we observe any clustering of these variations in specific chromosomal regions.

SNVs in Known Functional Elements of the Genome

To investigate functional relevance of SNVs found in iPSCs after induction and expansion, we next focused on SNVs in exons, especially those not present in dbSNP (build 132) database as reported previously. High-quality sequencing revealed six SNVs in BC1 iPSCs residing within the coding regions of six different genes (**Table 3-3, 3-4**). Three of them are non-synonymous (**Table 3-3, 3-4**). The paired sequencing data also revealed six SNVs in coding regions of six different genes in BCT1 iPSCs (derived from the same CD34+ cells); two of them are non-synonymous (**Table 3-3, 3-4**). Of the 1,058 SNVs between BC1 iPSCs and CD34+ cells, only 2 lie in a CpG island near a promoter, and none lie within the sno/microRNA regions in the UCSC's sno/microRNA track. For BCT1 iPSCs, there were no SNVs in the CpG islands but one in the micro RNA mir-124-2 (**Table 3-3**).

Blood Lineage Differentiation of iPSCs Generated from Human Blood Cells by Non-Integrating Plasmids

In addition to embryoid body (EB) formation for *in vitro* spontaneous differentiation (**Figure 2-10A**), I also conducted direct differentiation of BC1 iPSCs to hematopoietic lineages. Known numbers of dissociated BC1 iPSCs were deposited in round-bottomed, 96-well plates and then aggregated the cells into EBs by centrifugation. Cytokines that stimulate mesoderm and hematopoietic specification were supplemented at different time points. By day 10-12, single rounded and blood-like cells can be observed at the edge of EBs (**Figure 3-5A**). All single cells were harvested on day 18 and stained by a panel of blood

markers. Fluorescence-Activated Cell Sorting (FACS) analysis showed 78% of cells were CD45+ and 60% of cells were CD34/CD45 double positive, the former indicated properties of blood cells and the later showed the generation of blood progenitors (**Figure 3-5B**). Fifteen thousand cells were also taken for colony forming cell (CFC) assay in total 3 wells. After 2 weeks, various colonies were observed with different colony forming units (CFUs) (**Figure 3-6A, B**). The data show the iPSCs can differentiate to blood cells.

Some harvested single cells from EBs were put in suspension culture for erythroblast expansion and maturation. By day 10, cells were subject to FACS analysis and found 9% of cells were positive for erythroblast marker CD235a (glycophorin A) (**Figure 3-7A**). The data suggested that BC1 iPSCs possess the potential for erythroblast differentiation. In a parallel assay, my colleagues and I cultured suspension cells from BC1 EBs for megakaryocyte differentiation for 5 days. We can detect the expression of megakaryocyte markers CD41a and CD42a in differentiated cells, indicating they committed to megakaryocyte lineage (**Figure 3-7B**; unpublished data). Taken together, the data show that iPSCs derived from human postnatal blood cells by non-integrating plasmids can differentiate into multiple blood lineages.

Non-blood Differentiation of iPSCs Generated from Human Blood Cells by Non-Integrating Plasmids

My colleagues and I also tested BC1 cell line's differentiation potential to non-blood lineages. When we cultured BC1 iPSCs in a condition to promote

mesenchymal stem cell (MSC) differentiation, expression of surface markers CD73, CD271 CD256 and CD146 can be detected (**Figure 3-8A**). The differentiated cells can also form colony forming unit fibroblasts (CFU-F) (**Figure 3-8B**, left) or stained positive for Alizarin red which can identify calcium in osteogenic differentiation cells (**Figure 3-8B**, right).

Moreover, collaborators also showed BC1 can be successfully induced to Schwann cells, peripheral nervous system cells (**Figure 3-9**) (Liu et al., 2012) and astrocytes (**Figure 3-10**) (Shaltouki et al., 2013). These molecular and cellular results demonstrated that blood-derived iPSCs by episomal vectors have multi-lineage differentiation capacity.

Discussion

Using PCR (3 sets of specific primers) and Southern blot, I showed episomal vector DNA did not exist either as episomes or anywhere in the genome of the iPSCs analyzed. The whole genome sequencing analysis provided additional evidence all vector DNA has been lost. These results demonstrated that reprogramming human somatic cells does not require genomic integration or continuous presence of exogenous reprogramming factors and thus removes one obstacle to the clinical application of human iPSCs. In the future, it would be useful to build a method that can eliminate episomal vector DNA more efficiently than the passive expansion and dilution approach.

Interestingly, I noticed that all 5 of the iPSC lines derived from blood cells by non-integrating vectors we analyzed so far clustered together (and close to a

group of hESC lines), based on promoter DNA methylation signatures. They appear distinct from other iPSCs derived by integrating vectors from either blood CD34+ cells or MSCs/fibroblasts (**Figure 2-1**). A recent study found that mouse blood cell-derived iPSCs more closely resemble ESCs epigenetically and functionally than those derived from fibroblasts (Kim et al., 2010). Whether the virus-free and integration-free human iPSCs derived from blood cells and by the non-integrating vectors are superior in quality (closer to the best human ESCs) than previous iPSC lines derived by integrating vectors remains to be determined.

Increasing evidence suggests that both DNA replication-dependent and -independent mechanisms are involved in promoting reprogramming, which is fundamentally an epigenetic process (Hanna et al., 2009). My studies in generating iPSCs from unfractionated human blood MNCs depends on a culture condition to obtain a proliferating cell population that is neither T nor B cells, before plasmid-mediated reprogramming. I will work on finding a refined culture condition to shorten the culture period from current 10-14 days to 2-4 days. It is also of interest to determine if the serial delivery of synthetic mRNAs encoding 4-5 reprogramming factors or purified proteins could reprogram human blood MNCs more efficiently than human fibroblasts as reported.

The facile method described here that reprograms blood cells after minimal culture and using plasmids provides several important advantages over other virus-free and integration-free methods. First, one-time delivery of 1-2 plasmids that are much easier to produce and more stable than proteins or mRNAs makes the method more attractive or feasible for generation of clinical

grade iPSCs under current good manufacture practice (cGMP) conditions. Second, the capability to reprogram PB MNCs not only provides a readily available source of human cells for iPSC derivation (Yamanaka, 2010), but also shortens the cell culture time required to prepare and prime the target cells for reprogramming (now < 14 days from blood versus \geq 4 weeks from skin biopsies). We can harvest PB MNCs directly from patients and make cells become patient- or disease-specific iPSCs. Patient-specific iPSCs represent a valuable tool to study poorly understood mechanisms of hematological diseases. My combined method of using MNCs from a few milliliters of blood and non-integrating plasmids brings the iPSC derivation technology to a new level, and offers greater promise for the use of patient iPSCs for disease modeling and future clinical applications in regenerative medicine.

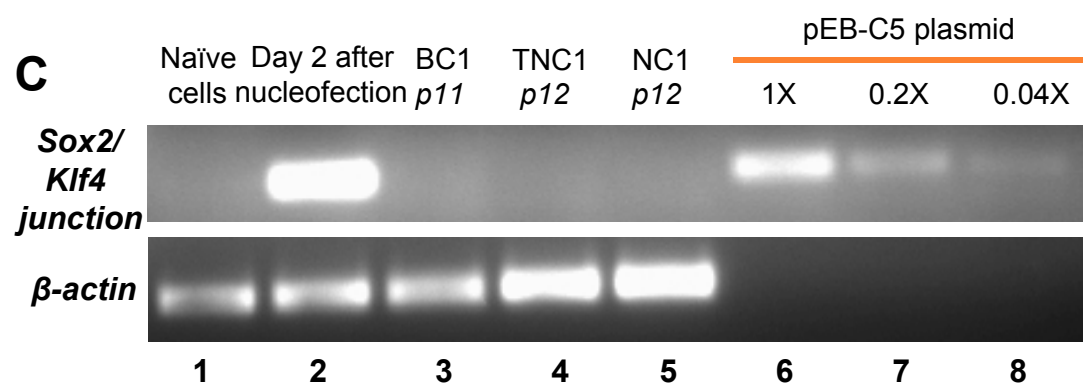
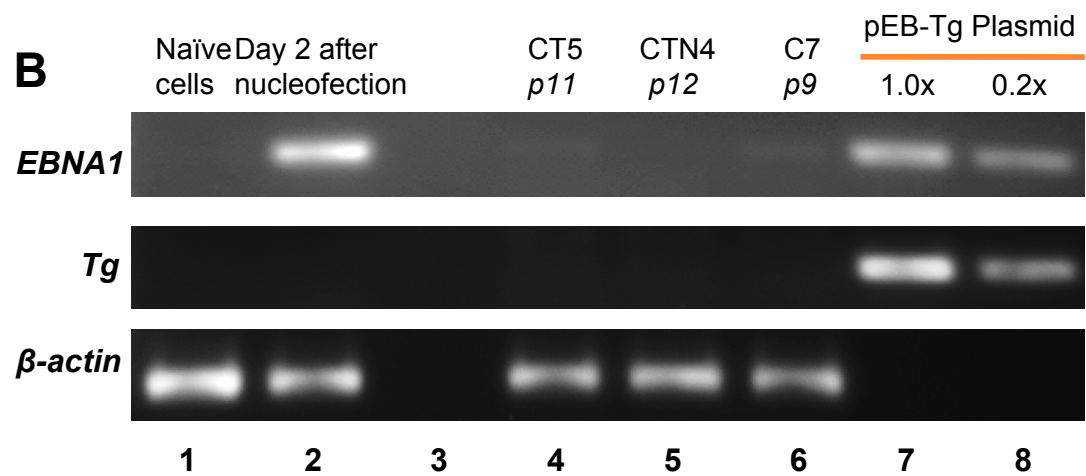
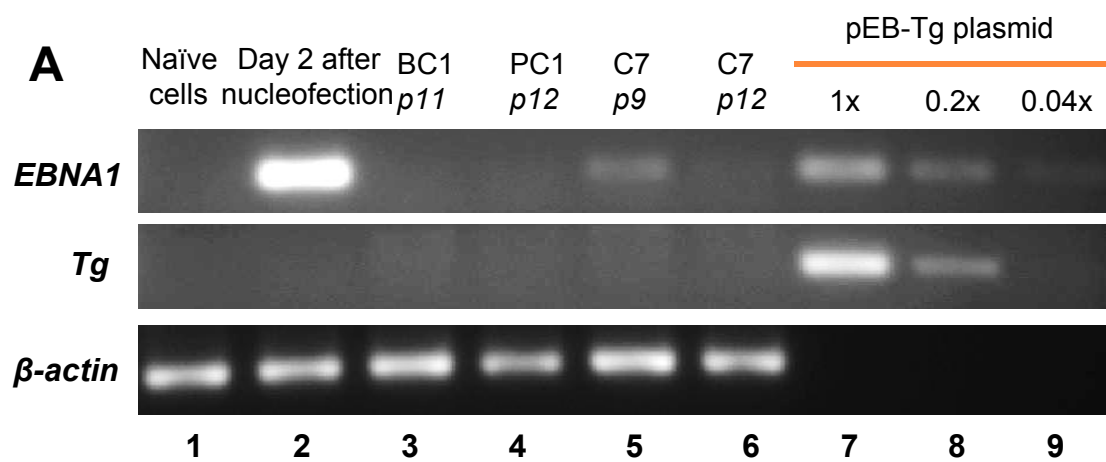


Figure 3-1. Detection of episomal vector sequence by PCR. (A) PCR-based detection of vector sequence EBNA1 and Tg in 3 iPSC lines derived from human various CD34+ cells by the single C5 vector. (C7: cord blood, BC1: bone marrow, PC1: mobilized peripheral blood). The single-copy cellular sequence β -actin was used as a quality control for isolated genomic DNA. Total DNA isolated from the cells before (naïve) and after nucleofection (day 2) were used as negative and positive controls for vector DNA. The pEB-Tg plasmid diluted in amounts equivalent to 1, 0.2 or 0.04 copies per genome of cellular DNA was used as a control for detection of the EBNA1 viral DNA sequence. **(B)** Two additional CB CD34+ cell derived iPSC lines by C5 together with Tg (CT5 and CTN4) were included. Genomic DNA was examined to detect the presence of EBNA1 or Tg sequence. **(C)** Another set of primers specific to Sox2/Klf4 junction sequence was used. TNC1 and NC1 iPSC lines were from PB MNC reprogrammed by C5+Tg or C5 alone.

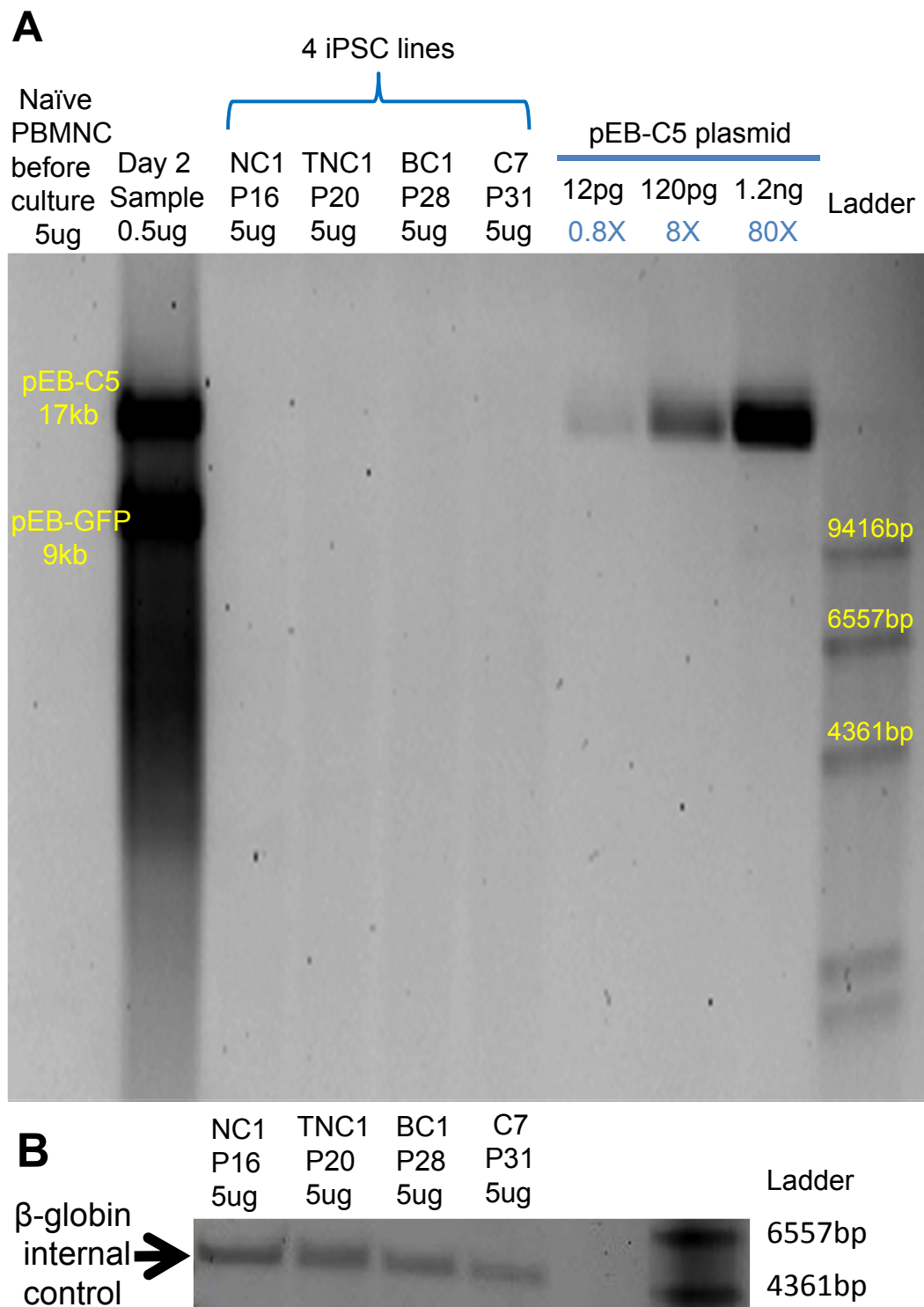


Figure 3-2. Southern blot analyses for the lack of vector DNA in expanded iPSCs that are derived by episomal vectors. (A) Detection of vector DNA sequences using EBNA1 DNA as a probe. Total DNA isolated from various cell extracts (lane 1-6) was digested by BamH I and separated by 0.7% agarose gel. Peripheral blood MNCs before (naïve, lane 1) and after nucleofection by two plasmids pEB-C5 and pEB-GFP at day 2 (lane 2) were included. Total DNA isolated from 4 iPSC lines were analyzed here: NC1 (passage 16, lane 3), TNC1 (passage 20, lane 4), BC1 (passage 28, lane 5) and C7 (passage 31, lane 6). 5 µg total DNA isolated from various cell types were used, except that only 0.5 µg DNA (10%) from the day 2 transfected cells was used in lane 2. Plasmid DNA of pEB-C5 was also cut (once) by BamH I, and used as a positive control in the detection (lane 7-9). Plasmid DNA was loaded in limiting dilutions (12 pg in lane 7, 120 pg in lane 8 and 1.2 ng in lane 9). The numbers of loaded vector DNA molecules per lane, relative to the copy numbers of nuclear genome of naïve cells and iPSCs, would be 0.8x, 8.0x and 80x, respectively. Ladder: DNA size markers. **(B).** Quality controls of genomic DNA isolated from 4 iPSC lines. The same genomic DNA samples were digested and analyzed by Southern blot. The probe used was a cellular DNA fragment 3' to the *HBB* gene. As expected, the method detected the 4.3-kb DNA fragment in iPSCs as a single-copy gene.

Table 3-1. List of characterized integration-free human iPSC lines from blood cells by 1-2 plasmids												
iPSC Lines	Cell Source	Episomal Vectors	± NaB	Karyotyping (passages tested)	Pluripotency Markers ¹	<i>in vitro</i> Pluripotency Assay ²	<i>in vivo</i> Pluripotency Assay ³	ESC-like Global DNA Methylation Signatures	Genetic Identity ⁴	Presence of Vectors: (by PCR)	Presence of Vectors: (by southern blot)	Presence of V(D)J DNA Recombination
CT5	CB CD34+ AilCells, Lot#100204	pEB-C5,Tg	No	Normal (46,XY) (p7)	Positive	Yes	Yes (1/1)	Yes	Confirmed	Absent	N/D	N/D
CTN4	CB CD34+ AilCells, Lot#100204	pEB-C5,Tg	Yes	Normal (46,XY) (p8)	Positive	Yes	Yes (1/1)	Yes	Confirmed	Absent	N/D	N/D
C7	CB CD34+ AilCells, Lot#CBC091022A	pEB-C5	No	Normal (46,XX) (p7)	Positive	Yes	Yes (1/1)	N/D	Not Applicable	Absent	Absent	N/D
BC1	ABM CD34+ AilCells, Lot# BM2426	pEB-C5	No	Normal (46,XY) (p11)	Positive	Yes	Yes (4/4)	Yes	Confirmed	Absent	Absent	N/D
PC1	mPB CD34+ AilCells, Lot# A1868B	pEB-C5	No	Normal (46,XY) (p12)	Positive	N/D	Yes (2/2)	N/D	N/D	Absent	N/D	N/D
NaB: Sodium Butyrate (added during the reprogramming); N/D: not done												
¹ Staining for a panel of pluripotency markers including OCT4, NANOG, TRA-1-60, SSEA4 and alkaline phosphatase;												
² By embryoid body formation assay showing the presence of various cell types derived from all the embryonic 3 germ layers;												
³ By teratoma formation assay (No. of positive/No. of samples tested);												
⁴ Confirmation of genetic identity to the corresponding parental cell type by PCR detecting variable short tandem repeats. High-density SNP assays were also used to authenticate CT5 and CTN4 iPSCs. Three SPE iPSCs lines derived from the SCD003 donor all contain the sickle mutation (A to T at codon 6) in both alleles of the beta-globin gene.												

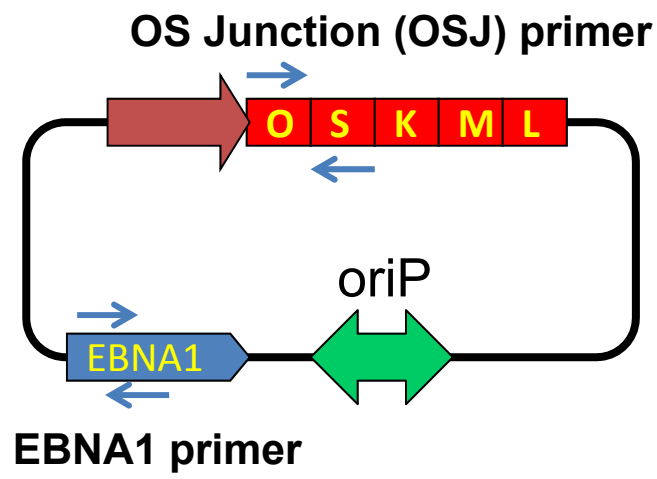
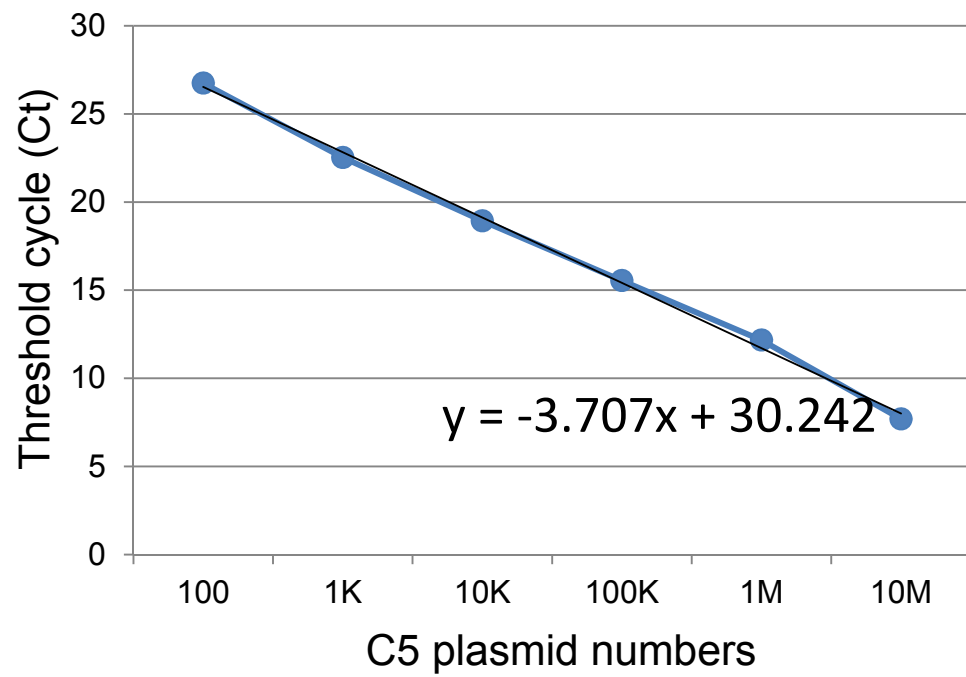
A**B**

Figure 3-3. QPCR to detect the dynamic change of episomal plasmid numbers in reprogrammed cells. (A) Two primers specific for EBNA-1 sequence or sequence crossing Oct4 and Sox2 transgenes were designated. **(B)** The threshold value of different amount of plasmid were derived to determine the standard curve.

Plasmid copy numbers per iPSC at different passages

iPSC passage number	Estimated plasmid copy #/cell	
	OSJ primer	EBNA1 primer
2 days after nucleofection	363.743	No data
p6	0.0025	0.0009
p8	0.0005	Undetectable
p9	Undetectable	0.0005
p11	Undetectable	0.0001
p12	0.0024	0.0019
p13	0.0001	0.0561
p14	0.0014	0.0004
p15	0.0210	0.0056
p16	0.0013	0.0004

Table 3-2. QPCR results of plasmid copy number per iPSC. 363 copies of plasmids per cell in the sample of 2 days after transfection. After passage 6 or higher, the plasmid number almost decreased to 0.

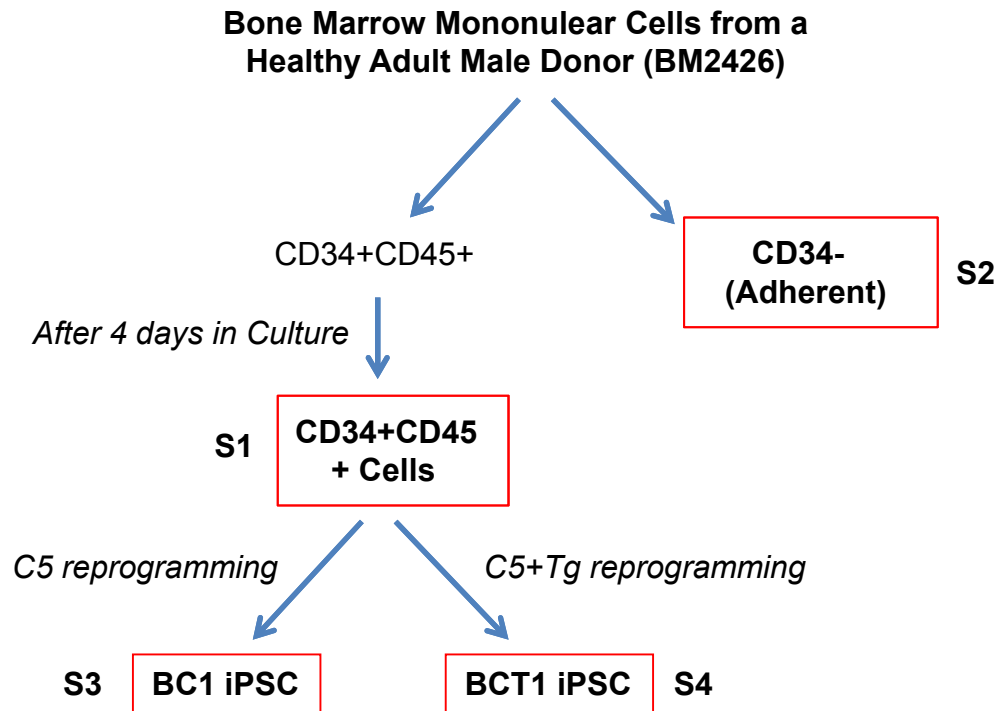


Figure 3-4. Relationship of iPSC lines and their parental somatic cells used in this study. Mononuclear cells from bone marrow (BM) of healthy adult donor (BM 2426) were separated into CD34⁺ cells (~1%) and CD34 depleted (CD34⁻) cells. The CD34⁺ cells were cultured for 4 days with hematopoietic cytokines before being reprogrammed by episomal vectors (left). The BC1 iPSC line was derived by using a single plasmid pEB-C5 whereas the BCT1 iPSC line was derived by addition of a second plasmid pEB-Tg. The BM CD34⁻ cells were used to establish cultures of marrow stromal cells. This was done in pair with the parental somatic cells (also boxed). BC1 and BCT1 iPSC lines were derived from the same batch of CD34⁺ cells cultured for 4 days.

Table 3-3. Summary of sequencing two pairs of iPSC lines and their parental somatic cells. No episomal vector sequence was found in iPSC lines BC1 and BCT1.

Features/iPSC lines	BC1	BCT1
Passages	25	18
Total nucleotides sequenced	172 Gb	153 Gb
Alignable nucleotides	165 Gb	142 Gb
Genome coverage (fold)	59X	51X
Total SNVs (compared to the hg19 reference)	4,158,672 (2,193,600)	4,207,199 (2,237,532)
Filtered single nucleotides variants (SNVs)	1058	1109
SNVs shared between BC1 and BCT1	No	No
SNVs cluster in specific chromosome region	No	No
SNVs selectively enriched in oncogenes	No	No
SNVs in coding regions	6	6
Synonymous (S)	3	4
Nonsynonymous (NS)	3	2
SNVs in CpG islands	2	0
SNVs in sno/microRNA regions	0	1

Table 3-4. Sequence variants in the coding regions in BC1 and BCT1

iPSCs

iPSC Lines	Gene Name	Chromosome	Position	Ref Allele	Variant Allele	Variant Type
BC1	CNBD1	8	88365930	G	A	nonsynonymous
	ITGA11	15	68650878	T	C	nonsynonymous
	SLIT2	4	20490512	C	T	nonsynonymous
	DHX34	19	47865827	C	T	synonymous
	DOCK1	10	128836023	C	T	synonymous
	PRPS1L1	7	18066668	A	G	synonymous
BCT1	ADAMTS17	15	100657067	C	T	nonsynonymous
	TG	8	133961089	C	T	nonsynonymous
	CDH4	20	60318830	G	A	synonymous
	EPPK1	8	144941146	G	C	synonymous
	MCM10	10	13213236	C	A	synonymous
	OR8K1	11	56113928	C	T	synonymous

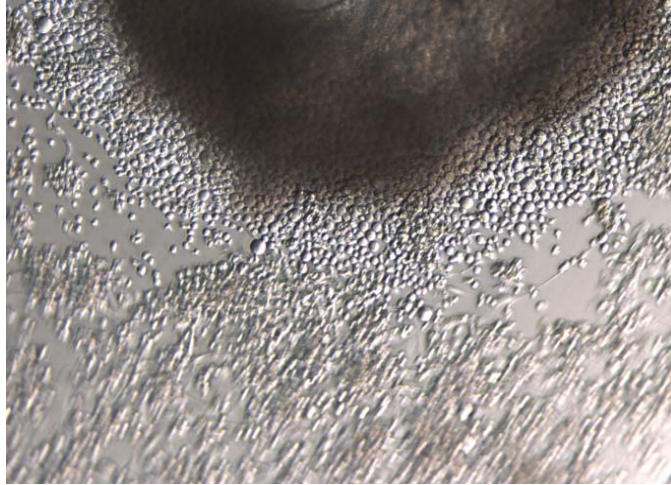
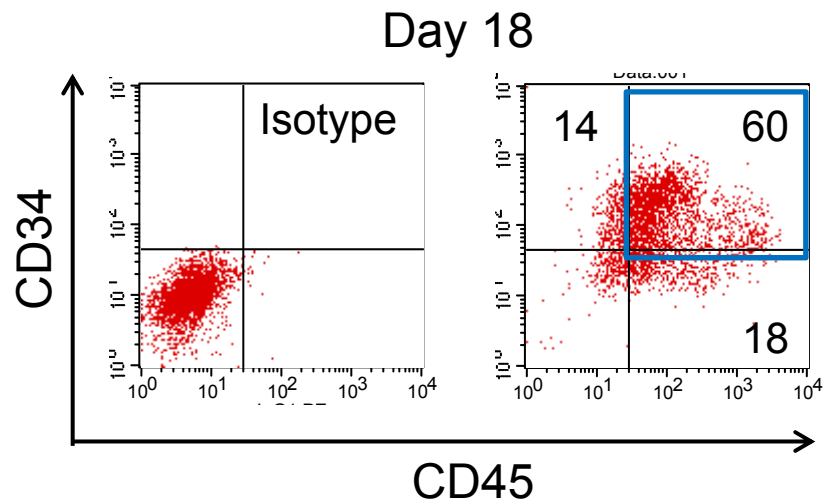
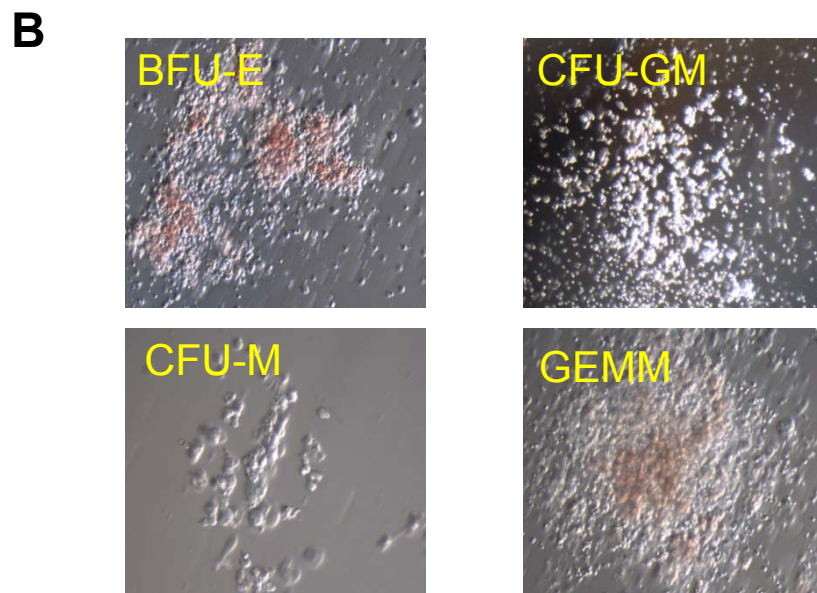
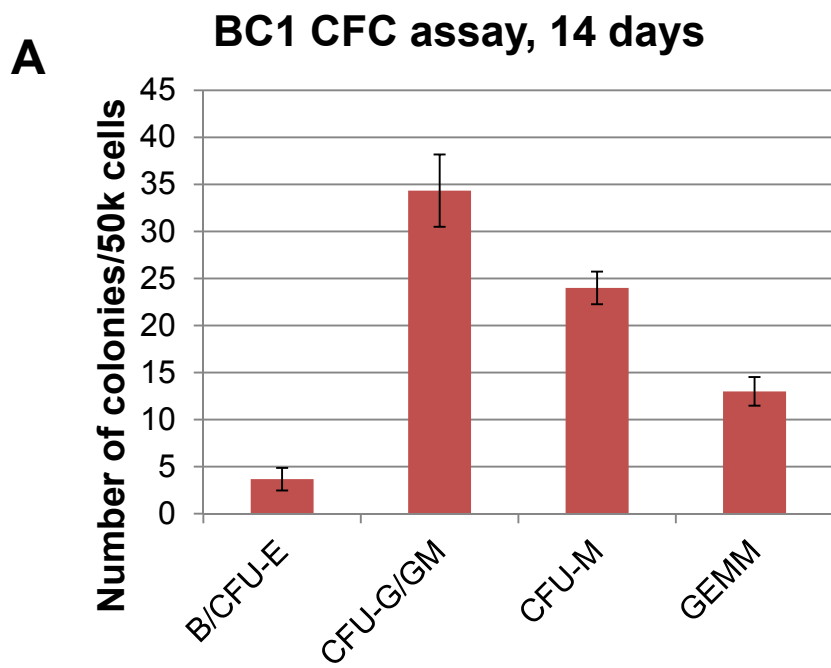
A**B**

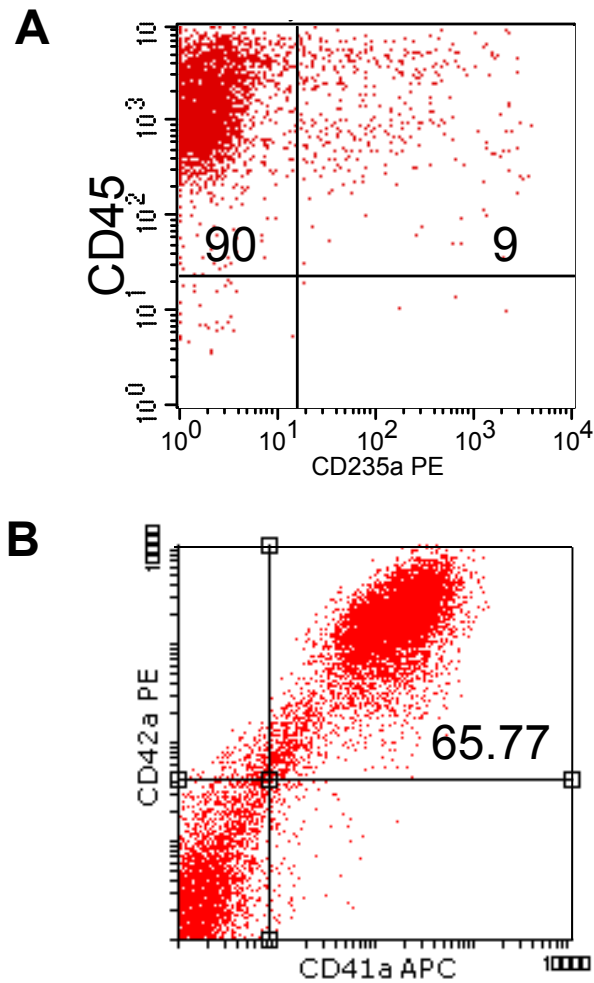
Figure 3-5. Morphology and phenotypes of BC1 iPSC derived hematopoietic cells. (A) Rounded hematopoietic cells emerged from EBs. **(B)** Emerged cells can express CD45 and CD34 markers. 60% of cells were CD34 and CD45 double positive, an indication of hematopoietic progenitor cells.



(Unpublished data, Wang, J)

Figure 3-6. CFC potential of BC1 iPSC derived hematopoietic cells.

(A) BC1 derived hematopoietic cells by EB formation were plated to 3 methylcellulose plates. Different colony forming units were identified and counted on day 14. **(B)** Human iPSCs derived from postnatal blood cells readily gave rise to multiple blood cells including erythroid progenitors. 10X magnification. Abbreviations: CFU-E (colony forming unit-erythroid); CFU-GM (colony forming unit-granulocyte/macrophage); CFU-M (colony forming unit-macrophage); CFU-GEMM (Colony forming unit-granulocyte, erythrocyte, macrophage, megakaryocyte).

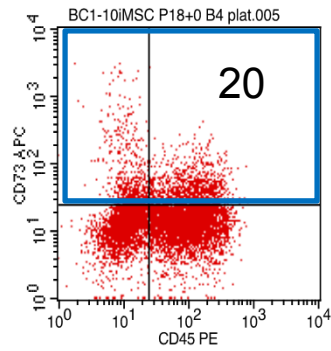


(Unpublished data, Gao, Y)

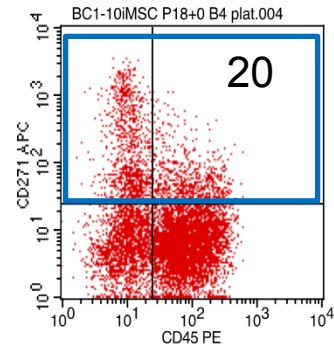
Figure 3-7. Erythroblast and megakaryocyte differentiation of BC1 iPSCs. **(A)** Surface antigen profiling by FACS in BC1 iPSC derived erythroblasts for erythroblast marker CD235a (glycophorin A). **(B)** BC1 iPSCs can differentiate into CD41a and CD42a double positive megakaryocytes.

A

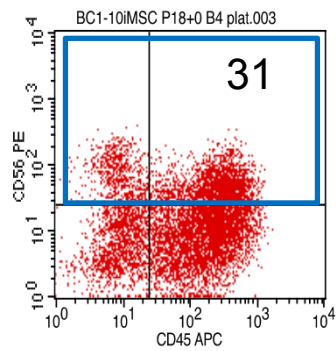
CD73



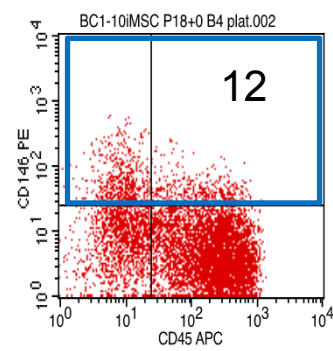
CD271



CD256

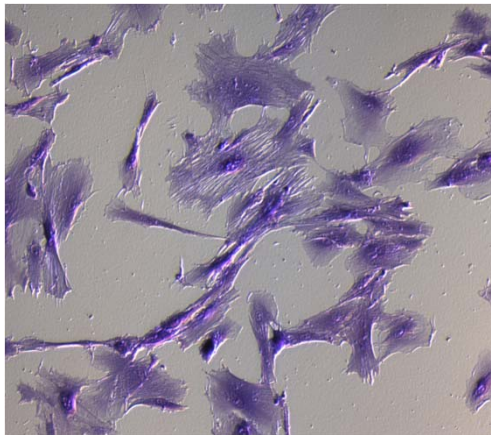


CD146

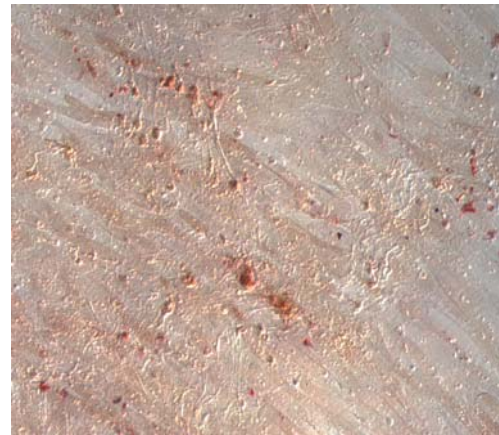


B

Colony forming unit fibroblasts
(CFU-F)



Osteogenic differentiation
Alizarin red staining



(Unpublished data, Tsang, K)

Figure 3-8. Mesenchymal stem cell differentiation of BC1 iPSCs. (A)

Surface antigen profiling by FACS in iPSC derived MSCs for CD73, CD271, CD256 and CD146. **(B)** Left: Morphology of BC1 derived MSCs stained by Crystal Violet-stained colonies. 20X magnification. Right: osteogenic differentiation by Alizarin Red staining. 20X magnification.

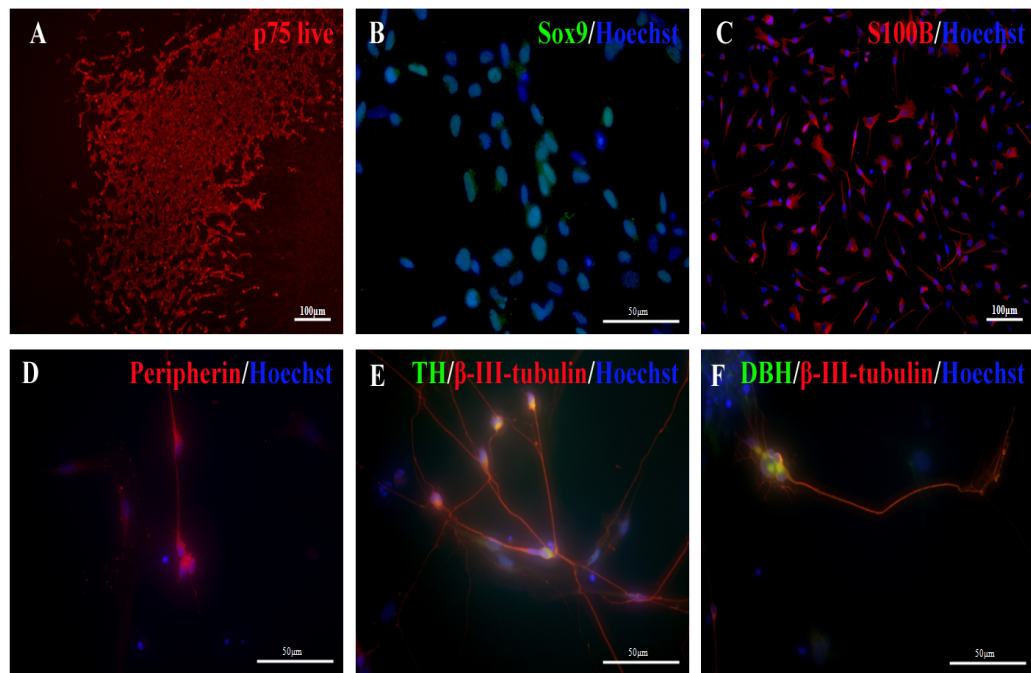
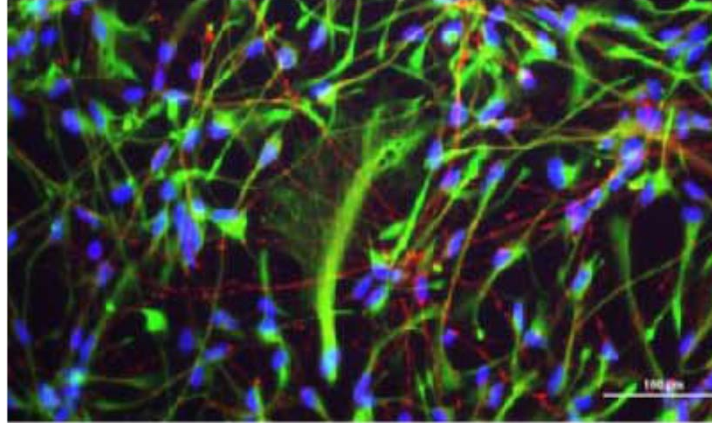


Figure 3-9. BC1 iPSCs can differentiate to Schwann cells and PNS neurons. (A) The cells emigrating out of the rosettes of BC1 iPSCs were live stained by p75 and could be purified by FACS. Cultured neural crest stem cell expressed Sox9 **(B)**. **(C-F)** BC1 iPSCs can differentiate into neural stem cells and give rise to S100 β ⁺ Schwann cells **(C)**, peripherin⁺ sensory neurons **(D)**, and sympathetic neurons positive for TH and a noradrenergic marker DBH **(E-F)**. Abbreviations: TH, tyrosine hydroxylase; DBH, dopamine-hydroxylase. PNS: peripheral nervous system. (Liu et al., 2012)

GFAP/Tuj/Hoechst



GFAP/S100β/Hoechst

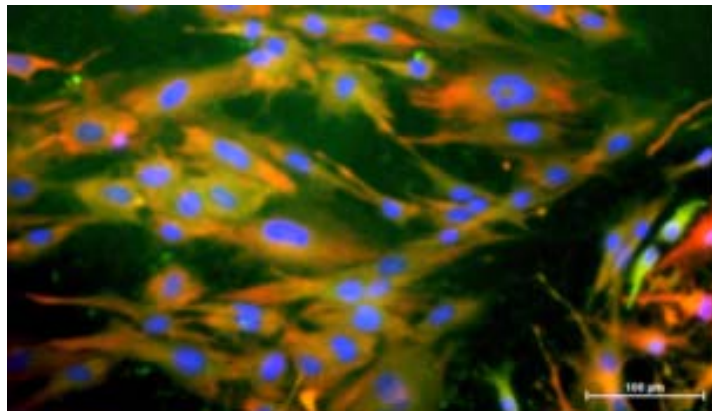


Figure 3-10. Differentiation of BC1 iPSCs into astrocytes. (A) BC1 differentiated for 5 weeks in defined medium displayed high percentage of glial fibrillary acidic protein (GFAP) immunostaining. **(B)** Cytoplasmic co-localization of S100 β and GFAP in 4 weeks differentiated BC1. Scale bars, 100 μ M (Shaltouki et al., 2013).

Chapter 4

Engineering Human Induced Pluripotent Stem Cells for GFP and Hematopoietic- Specific Reporter Systems

Background

Genetic manipulation of human pluripotent stem cells (hPSCs) provides a versatile tool to study stem cell differentiation, trace lineage commitment and may develop potential therapeutic applications. In the previous chapters, I have shown the evidence to derive human iPSCs from human blood cells via episomal vectors. In this chapter, I will address the application of genetically modified iPSCs for future differentiation tracking or lineage specification.

The green fluorescent protein (GFP) gene can be introduced and expressed in mammalian cells through transfection or virus infection. It has been widely used in cell and molecular biology as a biomarker (Tsien, 1998). Since iPSCs can proliferate indefinitely, labeling cells with long term expression of GFP would be useful to study the mechanisms governing differentiation. In the *in vivo* study, I can also detect GFP signal to locate the migration of iPSCs or determine cell origin.

Besides ubiquitous GFP expression iPSCs, generation of lineage-specific fluorescent protein reporters in human PSC lines also benefits researchers to investigate crucial mechanisms involved in lineage specification. Moreover, I can also isolate certain cell populations by cell sorting for detailed studies.

Wiskott-Aldrich Syndrome (WAS) is an X-linked recessive disease accompanied with thrombocytopenia, eczema and immunodeficiency (Massaad et al., 2013). It has been shown that WAS gene is only active within hematopoietic lineages including very immature cells (Parolini et al., 1997). Further studies also identified a short element of 170 nucleotides upstream of the WAS coding sequence regulates the expression of WAS gene (Petrella et al., 1998). These findings inspired me to engineer an iPSC line so that all blood cells differentiating from iPSCs can display GFP signal. This kind of reporter system will provide an exceptional opportunity to study how hematopoietic differentiation happens in stem cells and improve current differentiation protocols.

The RUNX1/AML1 gene is an important regulator for the development of definitive hematopoiesis in both mouse and human (Yamagata et al., 2005; Blyth et al., 2005). Loss of function of RUNX1 in murine embryo results in lethal phenotype due to lack of any definitive blood cells (North et al., 1999; Wang et al., 1996). However, the transcriptional regulation of this gene is largely unknown. In one previous study, researchers used comparative genomics and chromatin analysis to identify a conserved 531-base pair enhancer in the first intron of mouse RUNX1 gene (Nottingham et al., 2007). This enhancer plays an important role in RUNX1 gene expression. More strikingly, it targets reporter gene

expression to all known hematopoietic tissues of the mouse embryo which is similar to endogenous RUNX1 gene expression (Nottingham et al., 2007). In light of this result, I wonder if I can take advantage of this enhancer element to generate a reliable hematopoietic reporter system in iPSCs.

Approach

Genetic modification of human PSCs is not an easy task due to very low efficiency of delivery methods. Currently, gene transfer into human PSCs is largely based on the use of viral vectors. In the meanwhile, the transgenes in human genome are susceptible to transcription-dependent silencing. Thus finding an efficient transfection method is the key toward generating a faithful reporter system in human PSCs.

The *piggyBac* (PB) transposon is a mobile genetic element that efficiently jumps between vectors and chromosomes through a "cut and paste" mechanism. In contrast to retro- or lenti- virus-mediated genome integration which may activate gene expression around the integration sites, the PB transposon causes a relative stable integration in the human genome (Wilber et al., 2007; Wu et al., 2006). PB not only efficiently mediates gene transfer into the mammalian cells but also can be precisely excised, leaving no trace sequence behind.

In the first part of my goal, I will use an optimized PB transposon system to mediate gene transfer for GFP expression under the control of a ubiquitous promoter to one of iPSC line. Next, I will combine the WAS promoter fragment

and the RUNX1 enhancer element to test the idea of creating a hematopoietic reporter iPSC line.

Results

Generation of Stable GFP Expression iPSC Line by *piggyBac* Transposon System

To test PB delivery experiments, I selected human iPSC line BC1 which was reprogrammed by single pEB-C5 vector from healthy adult bone marrow CD34+ cells (id=BM2426) as the model. I have demonstrated the pluripotency and normal karyotype of this cell line. I also confirmed there is no pEB-C5 vector sequence left in the genome, demonstrating it is a footprint-free authentic iPSC line.

The standard PB system can be divided into a donor plasmid that contains the transposon and a helper plasmid that expresses the transposase (Cary et al., 1989; Wilson et al., 2007). I constructed the GFP gene under the control of CAG promoter in PB vector. BC1 iPSCs were cultured as previously described (Chou et al., 2011) and 1 million cells were harvested and dissociated to single cells. Equal amount of transposon and transposase were transfected through nucleofection into BC1 cells (see Methods). GFP+ cells can be detected 48 hours after nucleofection (**Figure 4-1A**). I performed puromycin selection for 7 days to allow successful integrated and stable GFP+ cells to grow. Next I used cell sorting to collect the brightest cells and reseeded them back to standard iPSC culture in order to pick up individual clones (**Figure 4-1B**).

One representative clone (GFP-BC1) is shown (**Figure 4-1C**). This clone has been cultured for 67 passages and with no sign of decreased GFP percentage or intensity. Taken together, I derived a stable GFP expression iPSC line that can be beneficial for future *in vivo* or lineage tracking studies.

Generate RUNX1-Enhancer/WAS-Promoter Driven *piggyBac* Vector as the Hematopoietic-Specific Reporter

The original WAS 170-bp promoter was in lentiviral vector to drive a downstream GFP gene (leuci et al., 2009). I first transfected this construct to k562 cell line and confirmed the strong expression of GFP signal in hematopoietic cells (**Figure 4-2B**). However, the original LV vector did not contain any selection gene, so I sought to clone the hematopoietic-specific reporter fragment to PB vector. WAS-GFP fragment was PCR amplified in order to create new restriction enzyme cutting sites and then inserted into empty PB vector (**Figure 4-2A**, left). In addition, I also wanted to test if including the RUNX1 enhancer to WAS-GFP-PB construct can have additive effect for labeling hematopoietic cells. The RUNX1 enhancer element was obtained by PCR amplification from human iPSCs and subsequently cloned to TOPO vector. Then the enhancer element was separated from TOPO vector and cloned upstream of WAS-GFP fragment, finishing the construct RUNX1-WAS-GFP-PB (**Figure 4-2A**, right).

Specific Marking of iPSC-Derived Hematopoietic Lineage by the RUNX1-Enhancer/WAS-Promoter Driven *piggyBac* Vector

The indefinite expansion potential of human iPSCs allows for clonal selection of individual transgene-modified clones. Therefore, I sought to nucleofect the two hematopoietic-specific reporter constructs: WAS-GFP-PB and RUNX1-WAS-GFP-PB to human iPSC line BC1. A parallel CAG-GFP-PB construct was also used as efficiency and experimental controls. After nucleofection and antibiotic selection, individual clones were picked and expanded. I defined nucleofected cells as BC1 reporter iPSCs to distinguish from intact BC1 iPSCs.

I first confirmed BC1 reporter iPSCs did not express GFP signal (**Figure 4-2C**). The hematopoietic differentiation potential of BC1 derived from three different constructs was examined by an improved method of EB formation and differentiation under a feeder- and serum-free condition (Yu et al., 2009; Ng et al., 2005; Ng et al., 2008). I did not observe any GFP expression in EBs in the first several days of hematopoietic differentiation. Strikingly, when single cells displaying blood cell morphology started to emerge from BC1 reporter EBs from day 9, I sporadically detected some cells were coincident with GFP expression (**Figure 4-3A**).

To further identify the phenotype of these GFP⁺ cells, all single cells were harvested from day 10 EBs for surface marker staining and FACS analysis. Undifferentiated BC1 iPSCs were negative for pan-hematopoietic marker CD45 and progenitor marker CD34 (**Figure 4-4A**). However, about 63% of

differentiated cells showed CD45+ and 37% of cells were dual positive for CD45 and CD34 (**Figure 4-4B**), indicating the high efficiency of EB hematopoietic differentiation. In particular, I observed 35% of single cells were GFP+, indicating the reporter GFP gene was expressed in BC1-derived hematopoietic cells. Although not all CD45+ cells were GFP+, most GFP+ cells indeed showed CD45+ marker (**Figure 4-4B**). I did not observe obvious difference between WAS-GFP-PB and RUNX1-WAS-GFP-PB constructs. Taken together, our preliminary data showed the functionality of hematopoietic-specific reporter system. It may help us to improve our protocol for a more efficient blood differentiation or study hematopoiesis in EBs.

Discussion

Genetic manipulation of human iPSCs to establish a reporter system is crucial to study stem cell lineage specification. It requires the capability to efficiently express a reporter gene. In this chapter, I first used *piggyBac* vector to deliver the GFP gene to the iPSCs derived in previous chapters. After selection and colony picking, one clone was picked and it displayed stable GFP expression with high intensity. My data showed even after blood lineage differentiation, most differentiated cells maintained GFP expression, suggesting the fidelity of GFP reporter. This cell line could be helpful for *in vivo* differentiation tracing.

I also took one step further to generate the hematopoietic-lineage specific reporter. The GFP expression is under the control of the WAS promoter and the RUNX1 enhancer, both which were reported to express in hematopoietic lineage

cells exclusively. I first confirmed iPSCs bearing this reporter construct did not express GFP, but upon hematopoietic differentiation, GFP signal were expressed in iPSC-derived hematopoietic cells. Detailed FACS analysis showed most GFP+ cells co-localized with hematopoietic markers CD45/CD235a. However, only partial CD45+/CD235a+ cells were marked with GFP. To rule out the possibility that the BC1 reporter cell line originated from more than one clone, I dissociated the cell line to the single cell level for subclones. Subsequent differentiation assay showed GFP+ was still tagged in partial CD45+/CD235a+ cells. In addition, WAS/RUNx1 vector also marked a subpopulation of CD34+ cells. Whether GFP+CD34+ population can generate more CD45+ cells than GFP-CD34+ cells remains to be determined.

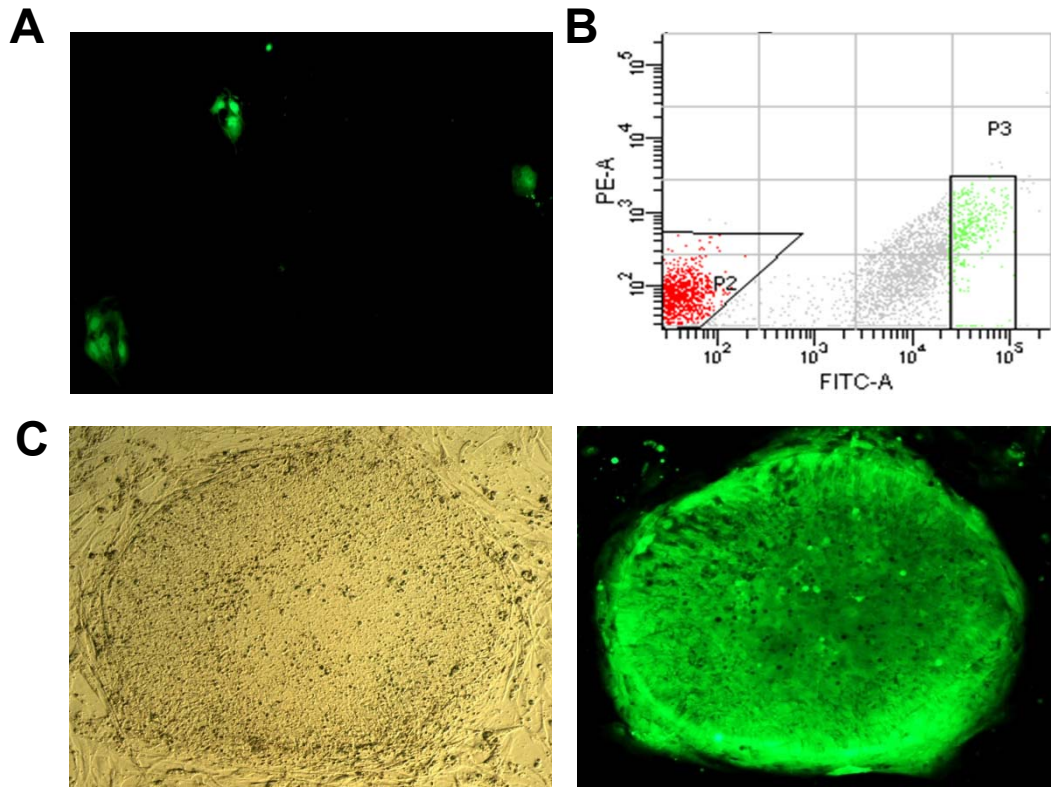


Figure 4-1 Genetic labeling of BC1 iPSCs with GFP via *piggyBac* transposon vector nucleofection. (A) GFP signal can be detected in BC1 iPSCs 48 hours after nucleofection. **(B)** Nucleofected cells dissociated to single cell level were collected by cell sorting for the brightest population. **(C)** Colony picking of single GFP+ clone after cell sorting and expansion. 10X magnification.

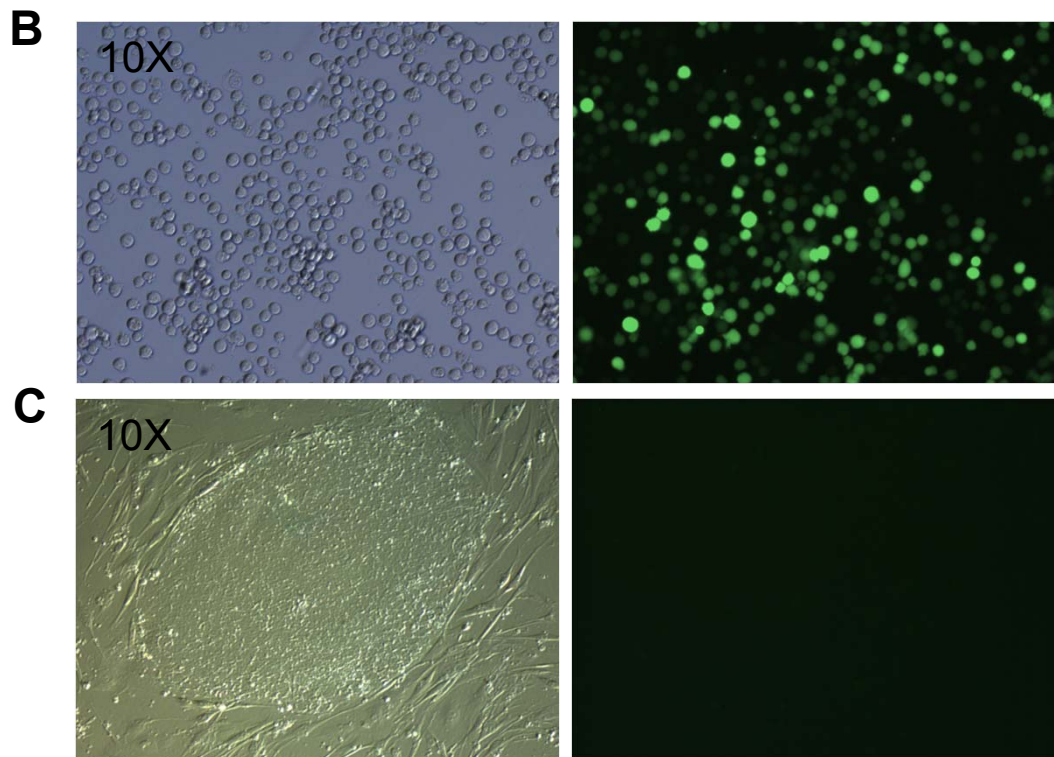
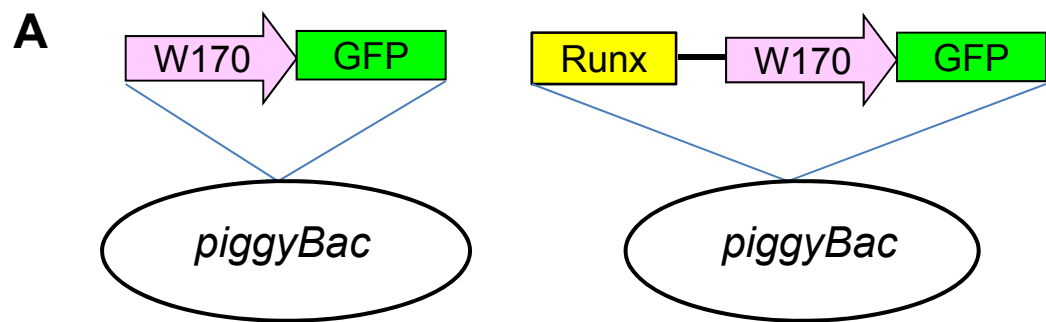


Figure 4-2. Expression of WAS-promoter driven *piggyBac* in blood cells. **(A)** Representation of the *piggyBac* vectors used to nucleofect human iPSCs. WAS-GFP-PB vector contains a 170-bp fragment of the human Wiskott-Aldrich Syndrome (WAS) gene proximal promoter driving the reporter gene GFP. In another set of reporter vector, a RUNX1 enhancer element was cloned to the upstream of WAS proximal promoter. **(B)** K562 cells infected by the original lentiviral vector containing WAS-GFP fragment showed GFP signal, proving the functionality of WAS-GFP cassette in hematopoietic cells. **(C)** GFP was not expressed in undifferentiated BC1 reporter iPSCs nucleofected with WAS-GFP-PB or RUNX1-WAS-GFP-PB vector. 10X magnification.

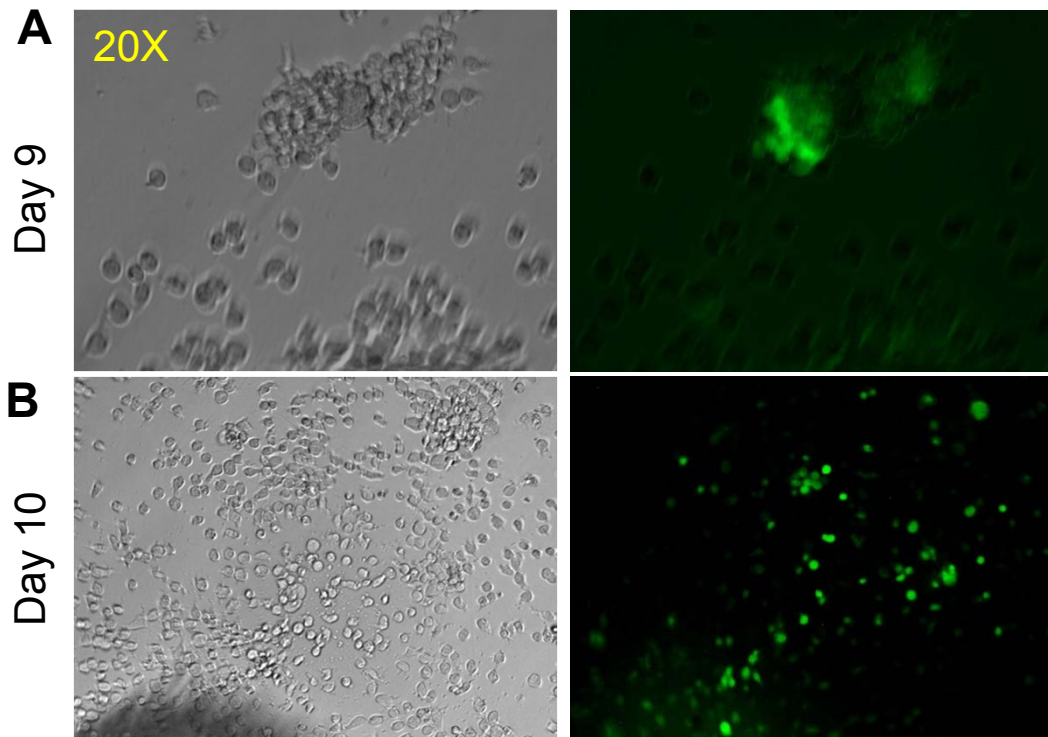
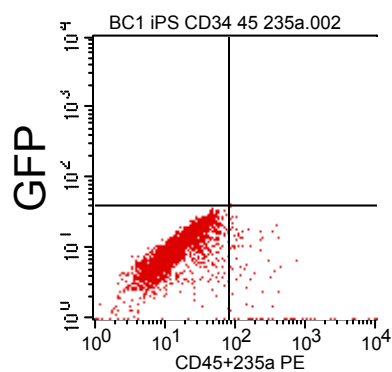
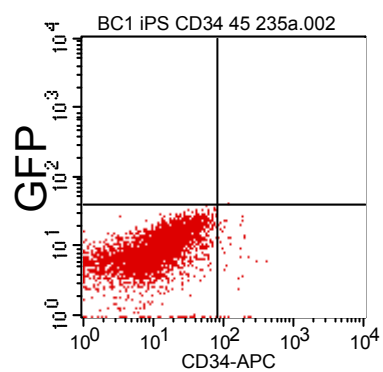


Figure 4-3. Specific expression of GFP in BC1 iPSC derived hematopoietic cells. (A) BC1 reporter iPSCs nucleofected by RUNX1-enhancer/WAS-promoter vector became GFP positive after EB formation for direct blood differentiation. **(B)** More blood-like single cells emerged from EBs and many gained GFP expression. Day 10. 20X magnification.

A

BC1 iPSC

**B**

BC1 reporter iPSCs

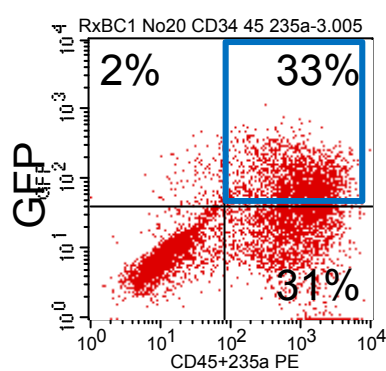
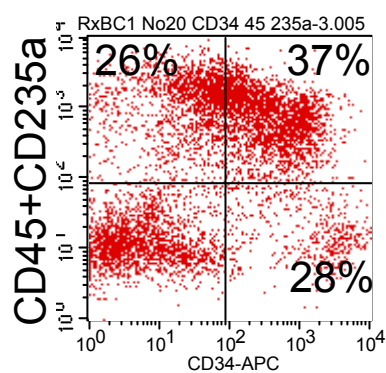
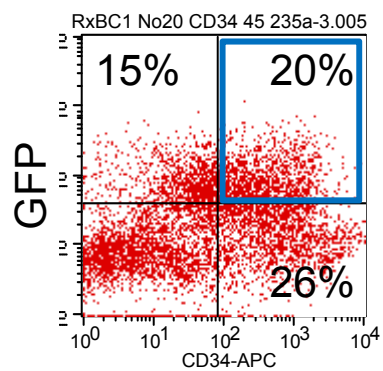
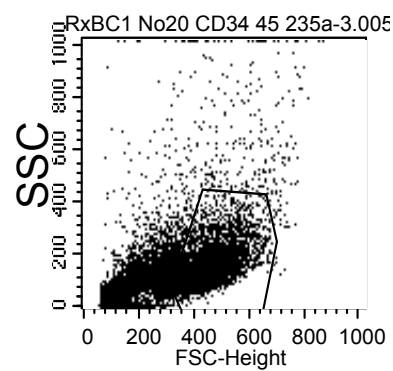
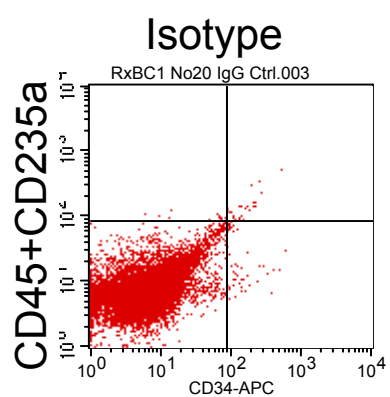


Figure 4-4. Hematopoietic differentiation induces GFP expression in RUNX1-WAS-PB and WAS-PB nucleofected BC1 reporter iPSCs. (A)

Undifferentiated BC1 iPSCs did not express blood-related markers: CD34 , CD45 or CD235a. BC1 iPSCs were also negative for GFP. **(B)** RUNX1/WAS-PB and WAS-PB nucleofected BC1 reporter iPSCs were prompt to blood differentiation and analyzed for GFP expression 10 days later. Isotype control was shown in the top row. Gated cells were subject for blood marker analysis (mid row left). Abundant cells expressed CD34 , CD45 or both markers, (mid row right). Some cells also showed GFP signal and partially co-localized with CD34 marker (framed, bottom row left). Not all CD45+ cells were GFP, but almost all GFP+ cells were coincident with CD45 marker (framed, bottom row right).

Chapter 5

Direct Reprogramming of Human Lineage Committed Blood Cells to Hematopoietic Stem/Progenitor Cells

Background

Stem cells hold great promise for cell therapy, especially hematopoietic stem/progenitor cells (HSPCs), the only stem cell type that has been widely used in bone marrow transplantation. HSPCs are a rare population and are defined by their ability to self-renew and give rise to all blood cell types. They are responsible for replenishing billions of blood cells every day. The classic sources of HSCs include umbilical cord blood (UCB), bone marrow (BM) and granulocyte-colony stimulating factor (GCSF) mobilized peripheral blood (mPB) (Yan et al., 1995). Although HSCs are a valuable model to study stem cells, the difficulty of expanding HSCs under defined, *in vitro* conditions impedes our understanding of hematopoiesis and also limits the potential clinical use of HSCs.

Efficient differentiation of pluripotent stem cells (PSCs) such as embryonic stem cells (ESCs) and iPSC into HSCs provides an alternative for generating HSCs and has been extensively pursued in the past three decades. Many *in vitro*

differentiation systems have been developed, either mediated by embryoid body formation or by co-culture with stromal cells lines such as OP9, a mesenchymal cell line derived from a mutated mouse strain with M-CSF deficiency (Nakano et al., 2009). Generation of various myeloid cell types from mouse and human PSCs, and even some lymphocytes at lower efficiencies, have been reported by many groups (Daley, 2003; Choi et al., 2009; Doulatov et al., 2013; Potocnik et al., 1994; Liang et al., 2013). However, the *in vitro* derived hematopoietic cells from PSCs typically resemble embryonic or primitive hematopoietic cells and lack robust *in vivo* engraftment and repopulation activities in conditioned recipient mice. Genetic manipulation such as overexpression of HoxB4 and other transcriptional factors in mouse ESCs can produce transplantable mouse HSCs (Kyba et al., 2002), although it is more skewed towards myeloid cells. However, this approach is less effective with human PSCs (Unger et al., 2008; Lee et al., 2008) as HoxB4 overexpression in human ESCs or even postnatal HSCs enhanced the growth of more differentiated myeloid progenitor cells instead of HSCs.

The studies of transcription factor (TF) -mediated cellular reprogramming or direct cell conversion question the long-held traditional concept that cell fate is not interchangeable (reviewed by Ladewig et al., 2013). By collecting lineage-specific TFs that are responsible for maintaining the target cells or inducing differentiation pathway toward the final products, many groups have reported trans-differentiation or direct conversion to neurons, cardiomyocytes and blood cells in mouse and human cells (Vierbuchen et al., 2010; Ieda et al., 2010; Qiang

et al., 2011; Pfisterer et al., 2011; Son et al., 2011; Marro et al., 2011; Pang et al., 2011; Caiazzo et al., 2011; Pereira et al., 2013). However, the process of selecting and optimizing the best lineage-specific TF combination for direct reprogramming is arduous and lengthy. It is unlikely that every time we can start from a pool of candidate genes, then narrow down to final essential factors. In an effort to devise a more general approach for dedifferentiation, a universal TF combination would be attractive.

In 2010, one group first showed the single OCT4 gene can directly reprogram human fibroblasts to multi-lineage blood progenitors (Szabo et al., 2010). Soon after, other groups demonstrated that the standard Yamanaka four factors can be used for direct reprogramming, instead of iPSC generation (Efe et al., 2011; Kim et al., 2011). In these reports, contractile cardiomyocytes or neural stem cells were derived. The underlying principle is that the four reprogramming factors bring cells to an epigenetically plastic and fate-undetermined intermediate (**Figure 5-1**). Providing appropriate developmental cues and culture conditions can guide cells toward a desired cell lineage. Reprogramming to pluripotency is just one of many possibilities after the four iPSC factor induction. Previous reports also have shown partially reprogrammed or intermediate stage cells displaying multiple lineage-specific markers (Mikkelsen et al., 2008; Meissner et al., 2007; Silva et al., 2008; Maherli et al., 2007; Sridharan et al., 2009; Thier et al., 2012; Matsui et al., 2012). These findings led me to ask whether the standard Yamanaka four factors Oct4, Sox2

Klf4 and c-Myc and appropriate culture conditions can directly reprogram differentiated somatic cells to HSPC.

Approach

To test the hypothesis that HSPCs can be directly reprogrammed from lineage committed cells, it is crucial to select an appropriate starting cell population. Human fibroblasts are the most studied cell type for direct reprogramming, but their epigenetic signature may be distant from the targeting cells--HSPCs. A cell type that is close to HSPC signature and easy to obtain and expand is ideal. Besides, an intra-lineage direct reprogramming may be more feasible to achieve. In previous chapters, I have developed a culture condition to expand various human blood mononuclear cells (MNCs) to erythroblast lineage (**Figure 5-2**). The expanded erythroid cells were positive for pan-hematopoietic marker CD45, indicating they remained the cells within the blood lineage. The exhibition of erythroblast markers after culture provides another advantage that they are erythroid lineage committed cells, and largely reduce HSPC contamination. I will focus on cord blood (CB) derived lineage committed cells because CB has higher growth potential and versatility than adult blood and may decrease the threshold of direct reprogramming.

Viral vector is still one of the most widely used methods to delivery transgenes to mammalian cells. But many viral vector-mediated reprogramming methods rely on multiple viral vectors to express individual reprogramming transgenes which result in multiple proviral copies throughout the genome and

the expression of each transgene is differential. Thus I decided to use a single retro viral vector expressing four reprogramming factors linked by 2A self-cleaving peptides to conduct direct reprogramming (Zhang et al., 2011). The polycistronic expression of four reprogramming factors from the same promoter should reduce the viruses necessary for reprogramming and make every time infection more consistent. In addition, the co-expression of GFP also helped me to track reprogrammed cells and served as an indication of transgene expression.

In addition, finding appropriate culture conditions to culture reprogrammed cells and support HSPC growth is important. Multiple growth factor combinations have been documented for supporting HSC expansion *in vitro* (Sauvageau et al., 2004). OP9 stromal cell co-culture systems and serum free culture conditions were also reported (Nakano et al., 2009). I will test variant culture conditions to expand potentially reprogrammed HSPC-like cells. Although the condition may not sustain the long term HSPC propagation, a short proliferation period should be enough to catch the cells for characterization.

Results

Characterization of Human Cord Blood Mononuclear Cells after Erythroblast/Megakaryocyte Expansion

Obtaining suitable blood cells as a starting population is critical for the direct reprogramming. It should remain in the blood lineage with capability to proliferate but lack HSPC features. Toward this end, I employed the method my colleagues developed for erythroblast lineage growth from human blood mononuclear cells (MNCs). Human cord blood (CB) MNCs were thawed and

cultured for erythroid cell expansion. During 10-14 days of culture, the cell number first dropped but rebounded around day 4 (**Figure 5-3A**). By day 10, cells kept growing and displayed morphology of typical blood cells (**Figure 5-3B**). FACS analysis confirmed most of cells remained pan-hematopoietic marker CD45+ (>99%) (**Figure 5-4**) and expressed erythroid progenitor marker CD36+ (>98%) (**Figure 5-4**), indicating the characteristics of erythroid-specific cells. My data showed ~34% of cells also expressed erythroid cell specific marker CD235a (glycophorin A) (**Figure 5-4**). In addition, only very few percentage of cells expressed CD34+ (<1%) and these minor cells will be excised by cell sorting (**Figure 5-4**). The expanded population contained rare TRA-1-60+ cells which is a surface marker for pluripotent cells (**Figure 5-4**). The signal was quite weak and could be just non-specific staining. Most cells were dim or negative for megakaryocyte markers CD41a and CD42a.

The expanded cells were first sorted based on CD45+ marker to exclude out any non-hematopoietic cells (**Figure 5-5A**). To ensure starting cells are free of HSPC contamination, only CD34- cells were sorted because HSPCs are enriched in CD34+ population. I also sorted cells expressing CD36 marker which is a common accepted marker for erythroid-specific cells (**Figure 5-5A**). The CD45+CD34-CD36+ triple selection provided a well-defined erythroid lineage committed cell population for following direct reprogramming assay.

To functionally characterize sorted cells, I seeded 5000 cells to methylcellulose media for colony forming cell (CFC) assay with 3 repeats. All colony forming units (CFUs) were counted on day 14 and revealed the majority of

BFU-Es and CFU-Es (**Figure 5-5B, C**). Taken together, these data proved an erythroid-lineage committed blood cell population raised from CB MNCs.

Generation of Cells with HSPC Traits via Retroviral Vector-Mediated Dedifferentiation

The sorted cells were subject to a polycistronic retroviral vector infection that expresses OCT4, SOX2, KLF4, c-MYC and GFP genes (**Figure 5-6A**). GFP enabled tracking of transgene expression during the reprogramming process (Zhang et al., 2011). In parallel, I also infected cells with a retroviral vector (Park et al., 2008) or a lentiviral vector expressing single OCT4 gene. A mock vector that expressed only GFP was used as an efficiency control.

Cells post-infection were plated to three distinct culture conditions for variant purposes. A portion of cells seeded on MEFs in the presence of NaB were for iPSC generation to monitor the quality of infection and cells. Some cells were cultured in serum- and feeder-free media (SFM). The remaining cells were co-cultured with OP9 stromal cells in serum-containing media. A combination of cytokines SCF, Flt3, TPO and IL-3 that favors the propagation of HSPCs were supplied to the latter two groups from day 3 post-infection throughout the experiment (**Figure 5-6B**).

Colonies with iPSC-like morphology were first observed between day 8 to day 12 in the control group, indicating the success of reprogramming. TRA-1-60 live staining also confirmed the genuineness of iPSCs (data not shown). Cells post-infection began to grow in OP9 free and OP9 co-culture conditions from day

14. GFP+ cells can be observed under the fluorescent microscope, indicating successful infection (**Figure 5-7**). Interestingly, some cells under OP9 cell co-culture displayed cobblestone morphology which could be a sign of HSPCs (reviewed by Bock, 1997) (**Figure 5-7**). In contrast, cells infected by GFP control vector did not show cobblestone phenotype (data not shown).

Growing cells from each culture condition were harvested for CD34 and CD45 marker detection on day 14 post-infection. Strikingly, I can detect a small population acquiring the CD34 marker (**Figure 5-8**) and maintaining CD45 expression in OP9-free and OP9 co-culture conditions. CD34 and CD45 double positive is usually an indication of HSPCs. This double positive population continued to increase only in OP9 co-culture condition (**Figure 5-6B, 5-8**). In contrast, CD34+CD45+ cells exhausted in the culture condition free of OP9. Prolonged culture showed these CD34+CD45+ cells can be kept for about 4-6 weeks with steady proliferation in OP9 co-culture condition.

To further analyze the property of growing cells, I harvest growing cells and sorted CD34+CD45+ cells for CFC assay. Total colonies were documented on day 14 (**Figure 5-9A** left). Interestingly, the sorted cells can generate various types of colonies such as CFU-E, CFU-G and CFU-M (**Figure 5-9A** right). This data revealed the CD34+CD45+ cells possess multi colony forming capability. Unexpectedly, CD34-CD45+ cell population can also form colony-forming progenitor cells, although with lower efficiency. Cells harvested from CFU study were subject to cytospin assay. Several blood cell types including macrophage, neutrophil-like cells and globin positive erythroblasts can be identified (**Figure 5-**

10). Taken together, these data provided evidence that multi-potent blood progenitors can be generated from lineage committed blood cells.

Discussion

In this chapter, I asked the feasibility of directly reprogramming cord blood derived lineage committed cells. This novel strategy may help us to derive valuable HSPCs which can't be expanded extensively *in vitro* or generated from PSCs. Based on the erythroid expansion culture condition described in chapter 2, I used expanded CB MNCs as the starting cell population. After infection of four iPSC-factors by retrovirus and cultured in the condition that favors the propagation of HSPCs, I observed a net growth of infected cells. Most strikingly, some of the growing cells displayed CD34 and CD45 markers which are often expressed in HSPCs. When CD34 and CD45 double positive were sorted for colony forming cell assay (CFC), multiple colonies were generated, demonstrating the *in vitro* differentiation capacity of reprogrammed cells. However, a thorough characterization such as *in vivo* functional studies and transplantation assays are required to verify this result.

By analogy with Waddington's epigenetic landscape, this result is similar to pushing differentiated cells uphill to a more immature state within the same valley (**Figure 1-1**, red outlined arrow), and can be depicted as dedifferentiation. Although dedifferentiation occurs naturally in living organism repair, special care should be taken in an attempt to do dedifferentiation in an artificial cell culture environment. First, there is still a chance that some cells capable of multiple-

blood-lineage forming reside in the CD45+CD34-CD36+ population. It should be vigorously examined that these cells are very rare or close to zero. Second, since a full set of four iPSC-factors was used for the infection, it is possible that some cells may be fully reprogrammed to iPSCs or go through a transient pluripotent state and then re-differentiate to CD34+CD45+ cells. To prove it is not the case, a culture condition that is not permissive for iPSC generation should be considered. For example, adding inhibitors such as oleate synthesis inhibitor can be an option to reduce the chance of full reprogramming (Ben-David et al., 2013). It would be also necessary to monitor cells at variant stages to check if pluripotent cells are generated. Direct single-cell reprogramming accompanied with time-lapse assay can help to demonstrate the occurrence of dedifferentiation. Eventually, reducing the transgene number and using factors not related to pluripotency are preferred.

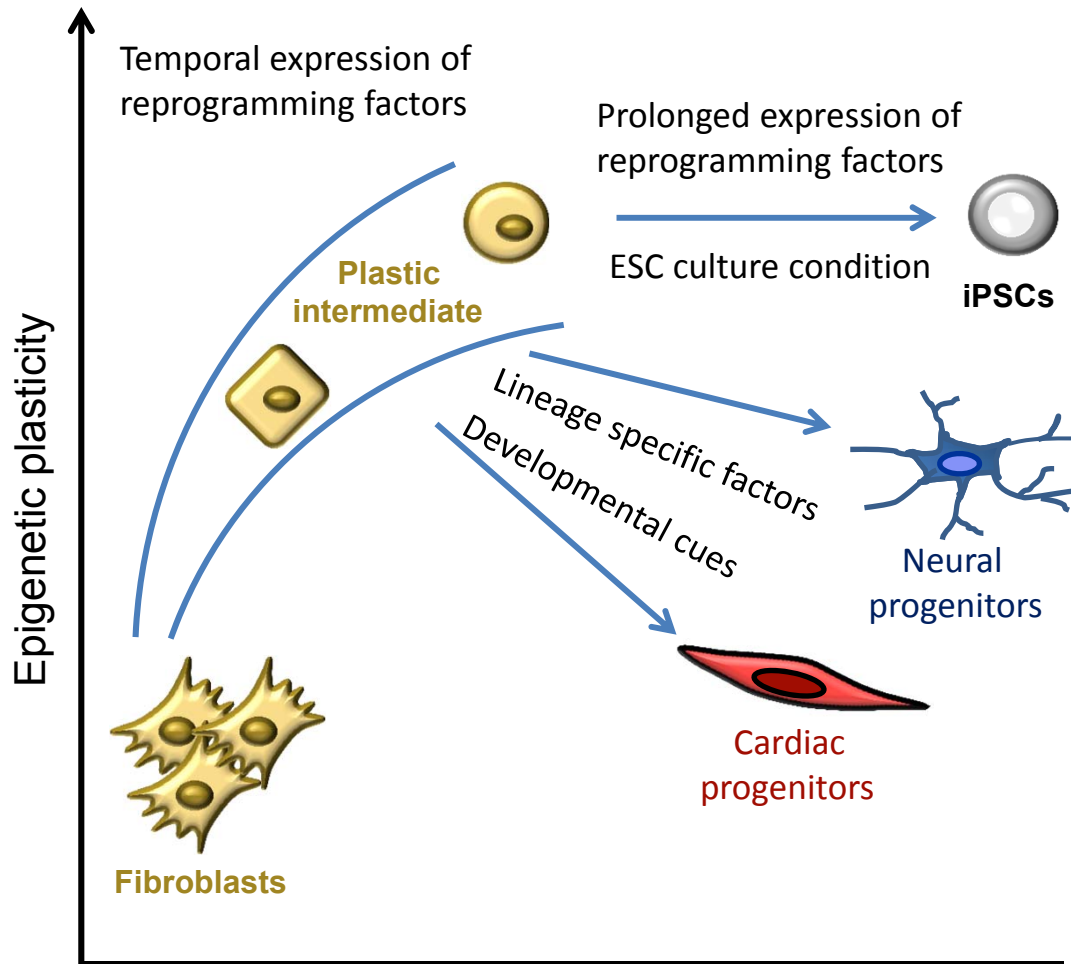
I noticed that once CD34+CD45+ cells emerged, they can expand in the presence of OP9 stromal cells for 8-10 weeks but then gradually decrease. Maybe the current culture condition is not ideal and can't support CD34+CD45+ cells to proliferate in a long term manner. Finding a better culture condition will be necessary and advance the result. Although this study is based on an assumption that lineage committed blood cells after four iPSC factor induction acquire a flexible epigenetic status, it is unknown whether current culture condition can induce authentic hematopoietic program to cells. Will other transgene combinations elicit direct reprogramming more efficiently? Are reprogrammed CD34+CD45+ cells bona fide HSPCs or just progenitor cells with

limited hematopoietic differentiation potential? I have shown these emerged CD34⁺CD45⁺ cells can give rise to various CFCs *in vitro*, indicating they have certain differentiation capacity. Whether these cells are capable of engrafting and repopulating to multiple lineage blood cells in immuno-deficient mice awaits more studies. Also the global gene expression analysis and epigenomic study can lend clues how they resemble authentic HSPCs.

In the meanwhile, it is noteworthy that the retroviral vector confers continuous expression of four iPSC transgenes in reprogrammed CD34⁺45⁺ cells. The prolonged expression of transgenes may hamper hematopoietic differentiation. Thus, switching to inducible or transient- expression system for direct reprogramming will help to turn off transgene expression and allow reprogrammed cells to achieve full hematopoietic differentiation.

In summary, the molecular and cellular data presented in the chapter showed multi potent blood progenitors can be derived from lineage committed blood cells via a direct reprogramming strategy. It provides the first evidence that it is possible to dedifferentiate blood cells to progenitor cells. In another study, it was shown a combined of transcription factors (TFs) Gata 2, Gfi1b, cFos and ETV6 can induce hemogenic program in mouse fibroblasts (Pereira et al., 2013). It will be interesting to test if human fibroblast can be converted to blood cells using the same combination or other factors. It is also important to compare whether the general TFs (OSKM) or a customized combination of TFs can generate blood progenitor cells that more closely resemble authentic HSPCs. The ultimate goal

in this study will be to generate therapeutically functional blood cells for clinical transplantation.



Adapted from Efe et al., *Nat. Cell. Biol.*, 2011

Figure 5-1. A model for direct reprogramming to alternative cell fates. Yamanaka 4 iPSC-factor induction may erase epigenetic state and bring somatic cells to a transient and plastic status. Providing ESC culture condition will induce cells to iPSCs, while exposing cells with distinct developmental cues or culture conditions may guide cells to specific lineages other than pluripotent state.

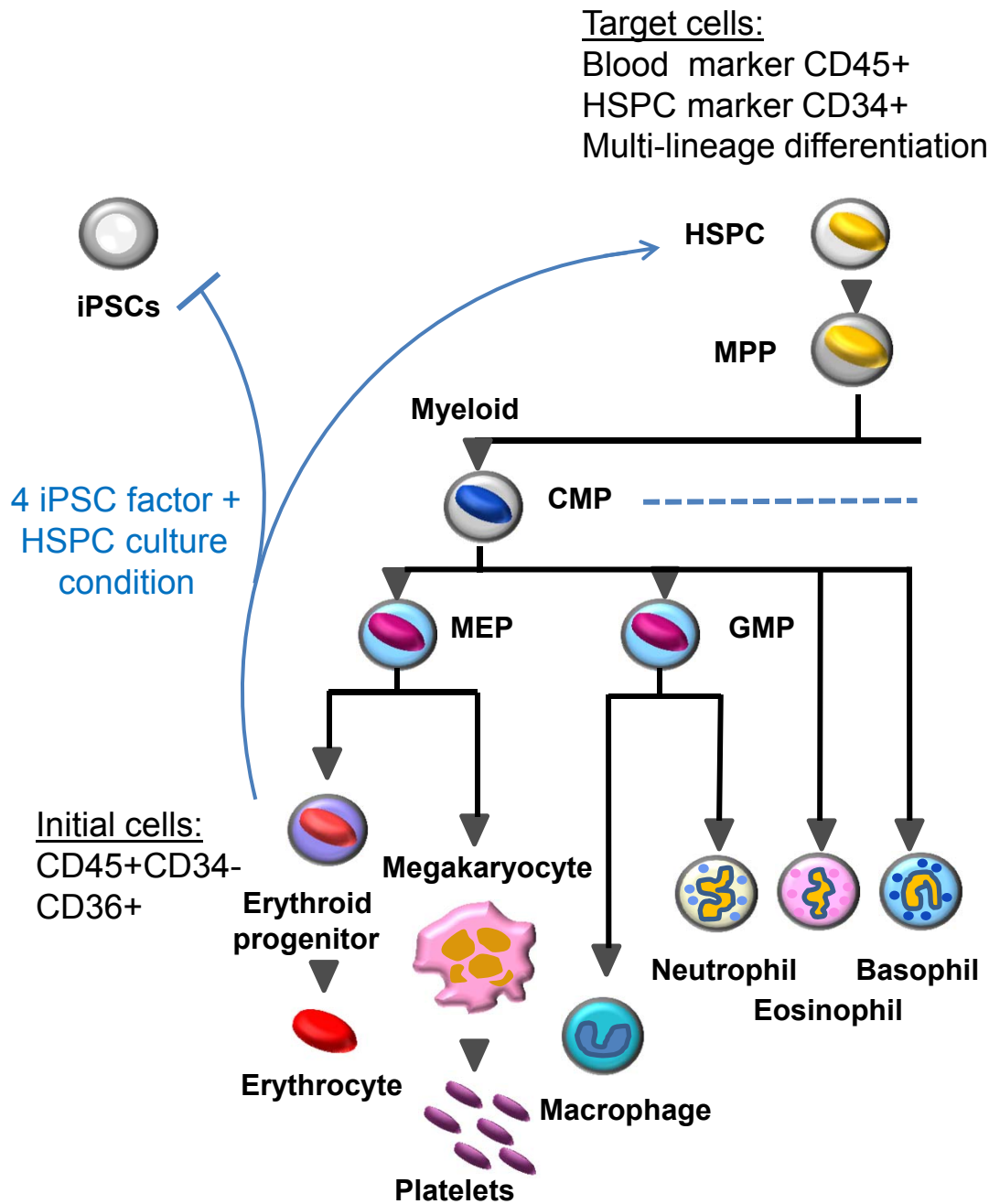


Figure 5-2. A scheme to show the hypothesis that lineage committed blood cells can be directly reprogrammed to more primitive blood cell types but not to pluripotency.

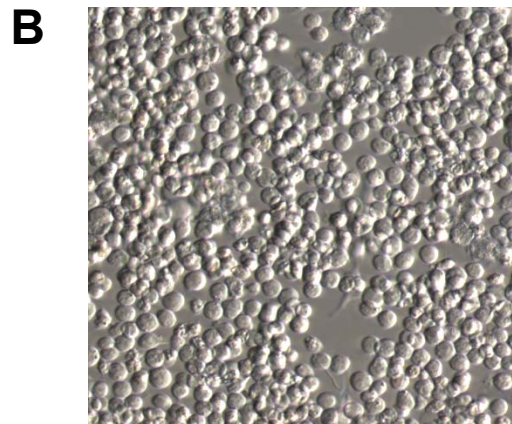
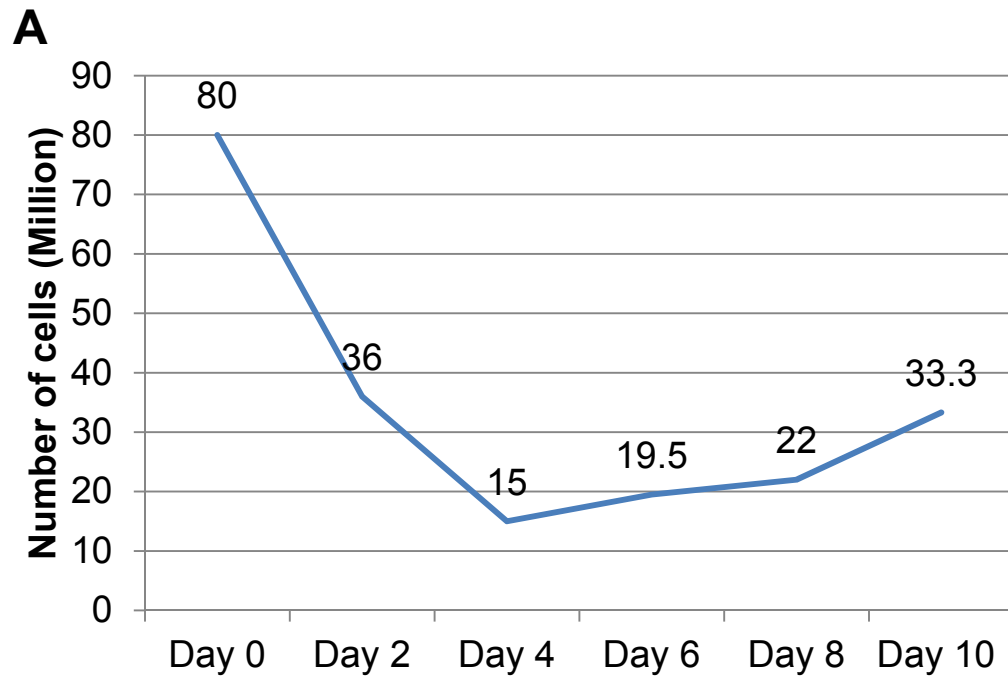


Figure 5-3. Human cord blood mononuclear cells (MNCs) under erythroid expansion culture condition. (A) Growth curve. (B) Cell morphology after 10 days of expansion. 20X magnification.

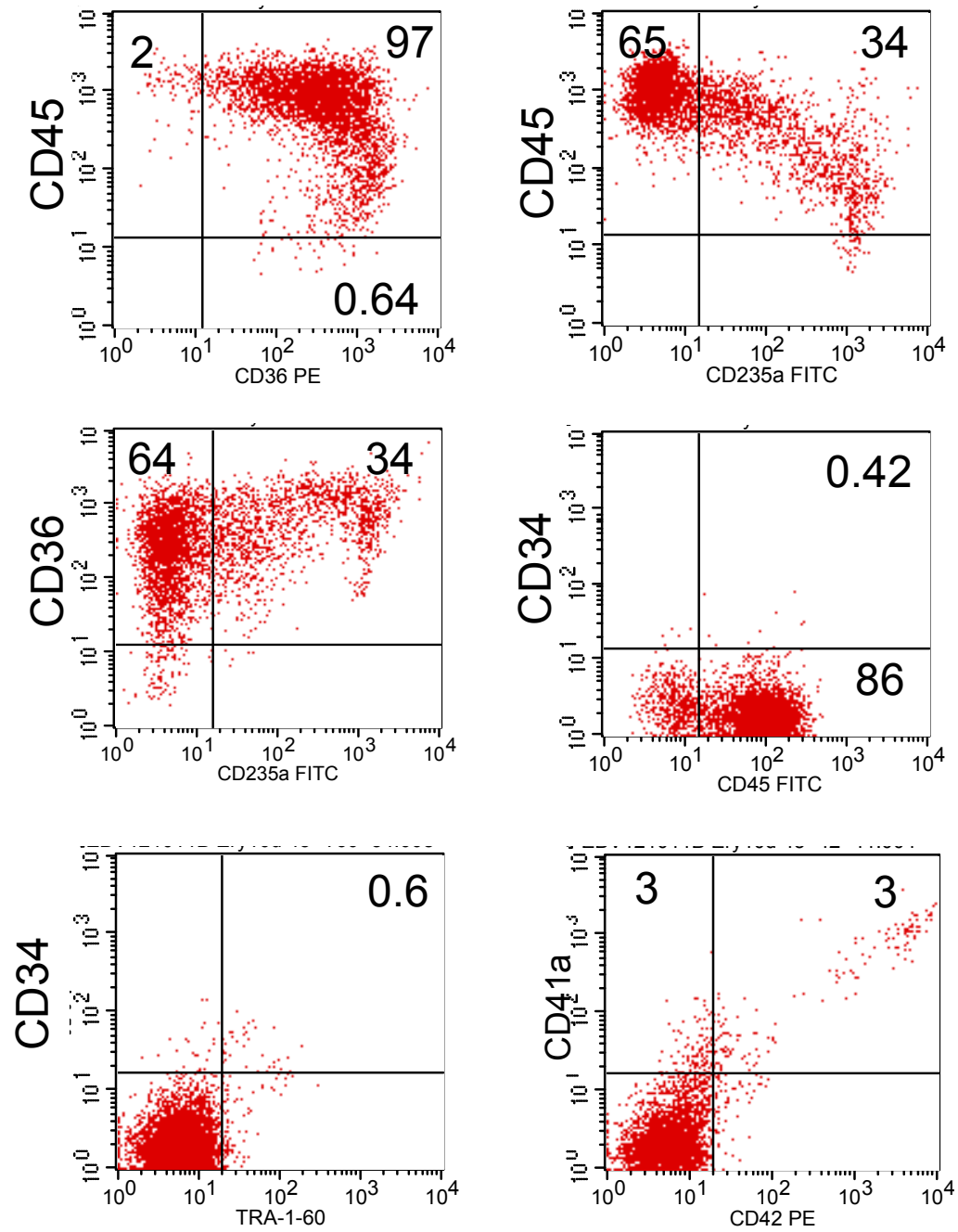
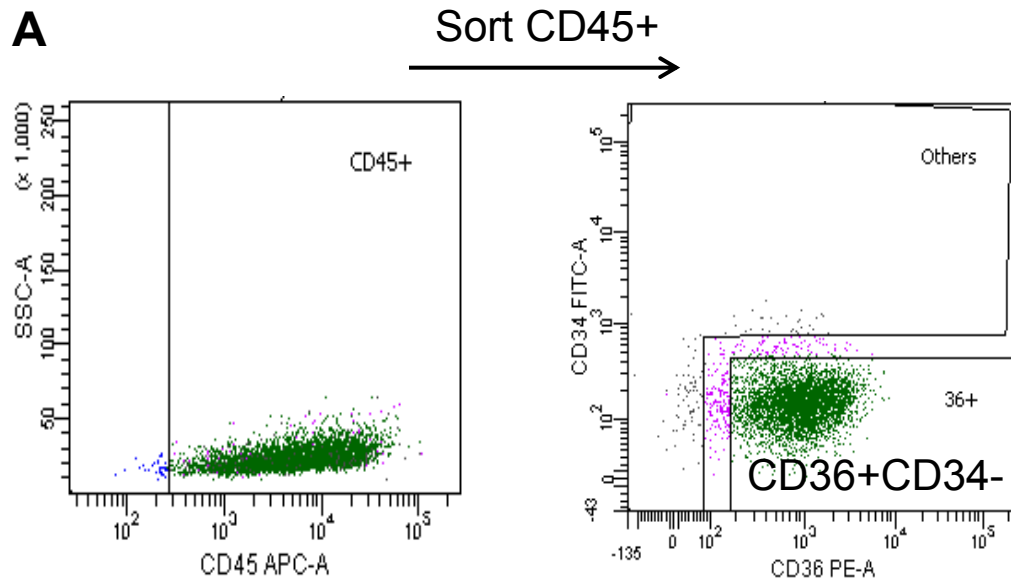
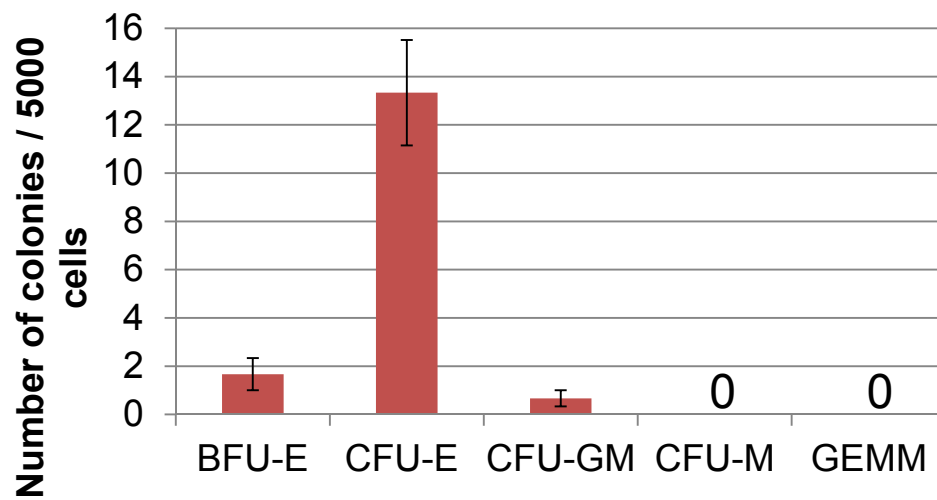


Figure 5-4. Characterization of expanded cells. Mononuclear cells expanded in erythroid culture conditions for 7-10 days expressed the pan-hematopoietic cell marker CD45 (99%) and the erythroblast markers CD36 (98%) and CD235a (34%), but not the hematopoietic stem/progenitor cell marker CD34 (<0.6%) or the pluripotent stem cell marker TRA-1-60.



B CFU assay of CD45+36+34- cells, 14 days



C CFU-E

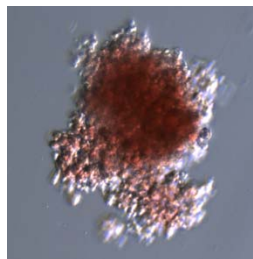


Figure 5-5. Selection of lineage committed blood cells for the direct reprogramming assay. (A) CD45+CD36+ but CD34- cells were sorted as a defined lineage-committed cell population and to exclude potential HSPC contamination. **(B)** Sorted cells mostly gave rise to erythroid colonies (CFU-E), confirming their erythroid lineage property. **(C)** A representative CFU-E colony.

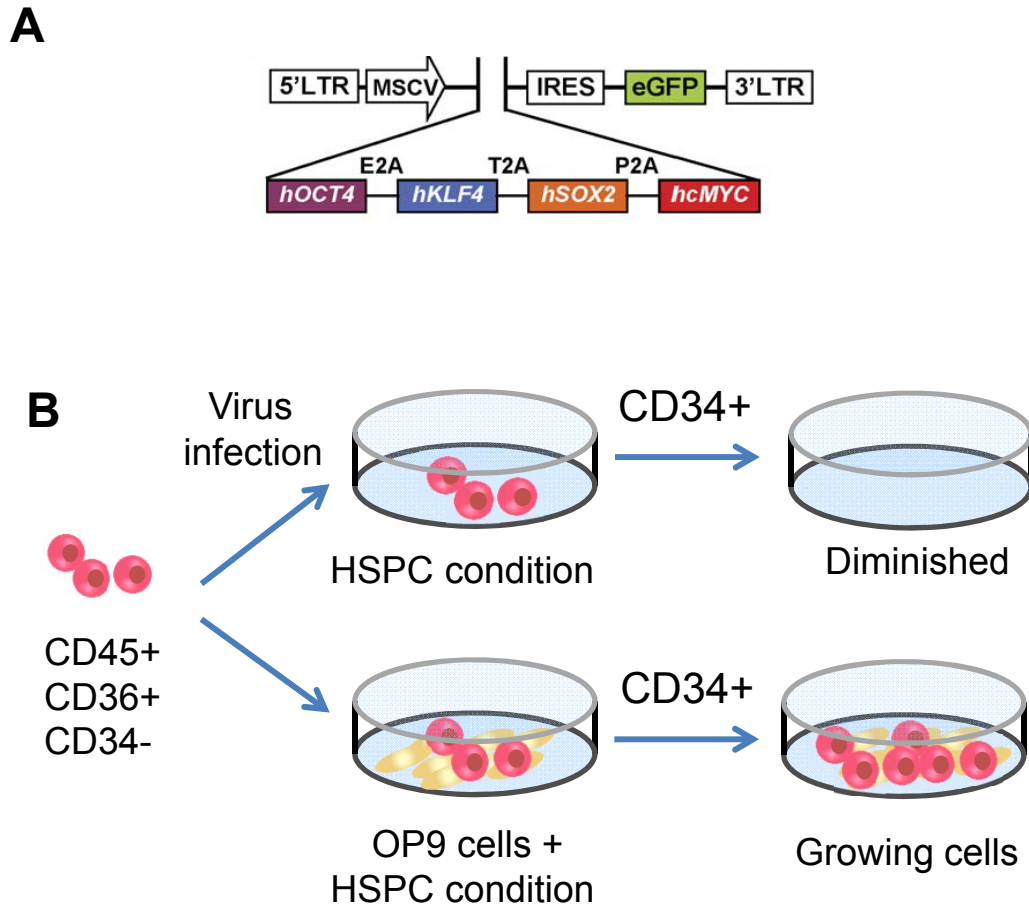


Figure 5-6. Infection and culture of reprogrammed blood cells. (A)

Schematic representation of the poly-cistronic retroviral vector expressing Yamanaka 4 iPSC factors and GFP (MIG4F; Zhang et al., 2011).

(B) After retroviral vector infection, cells were cultured in a HSPC condition in the absence or presence of OP9 stromal cell co-culture.

Only the OP9 cell co-culture condition supported the growth of emerged

CD34+ cells

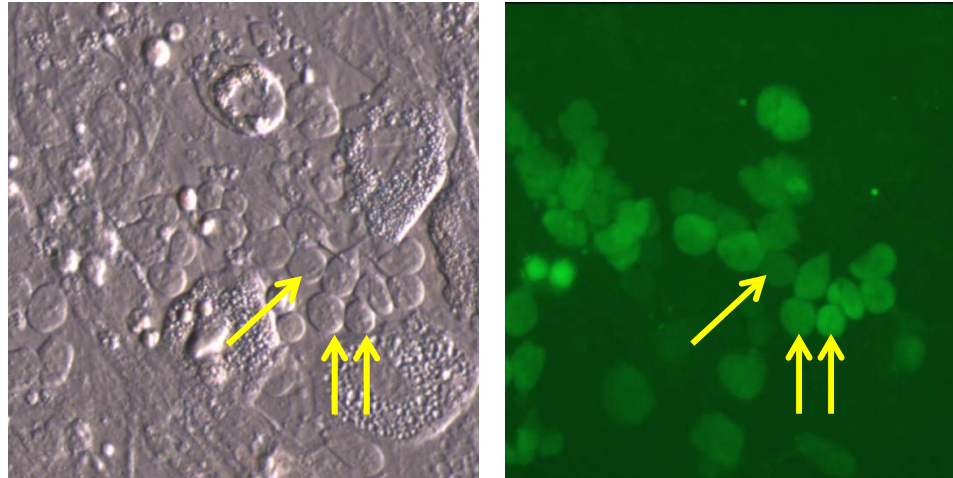


Figure 5-7. Lineage committed mononuclear cells successfully infected by the retroviral vector expressed GFP. Cobblestone-like blood cells can be observed in the OP9 co-culture condition, yellow arrows. Day 10 post-infection.

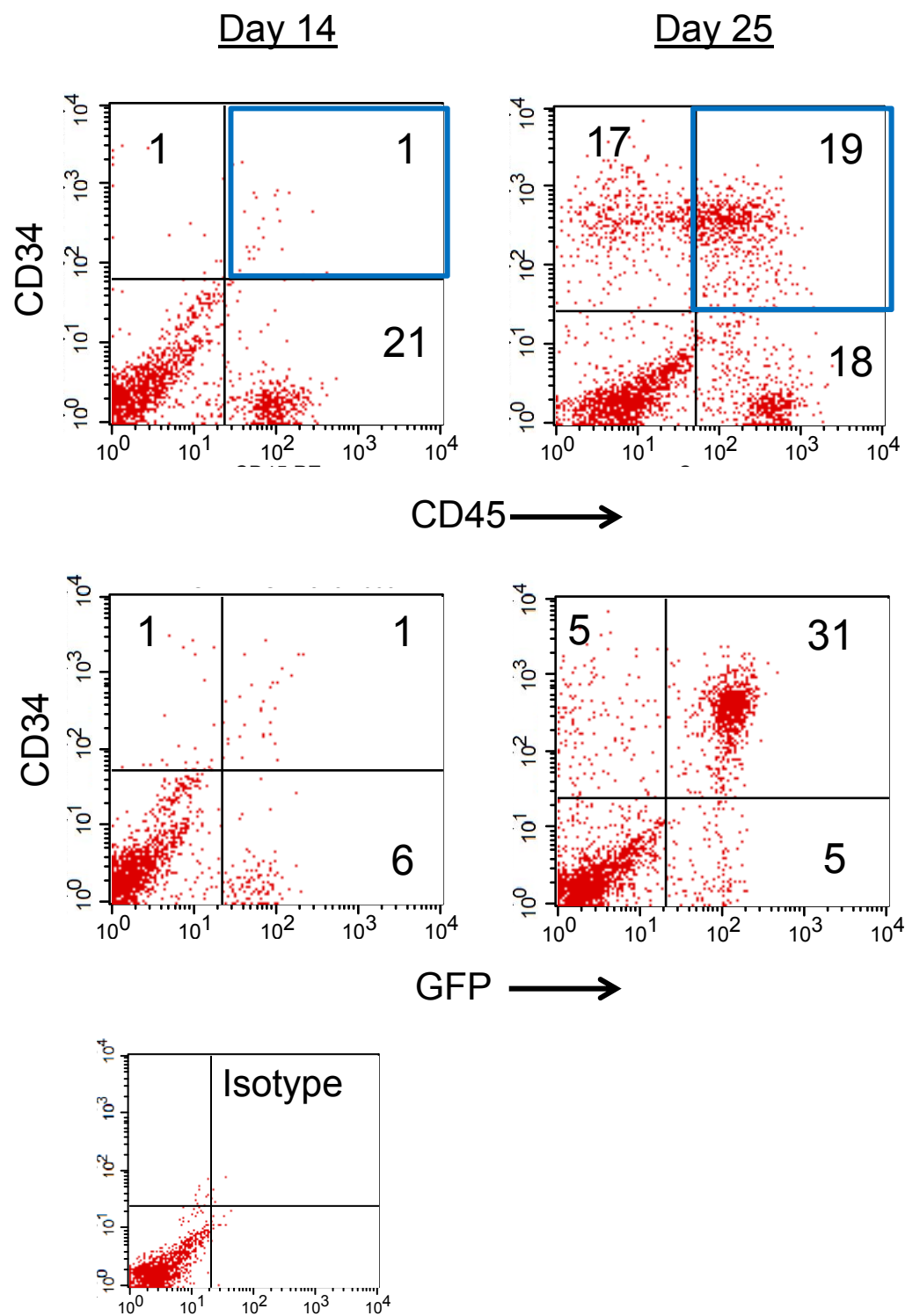
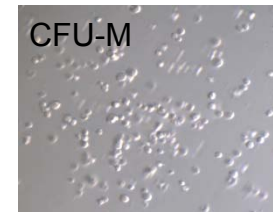
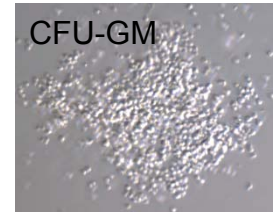
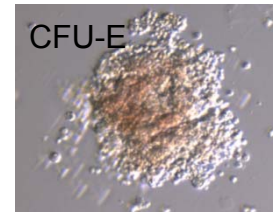
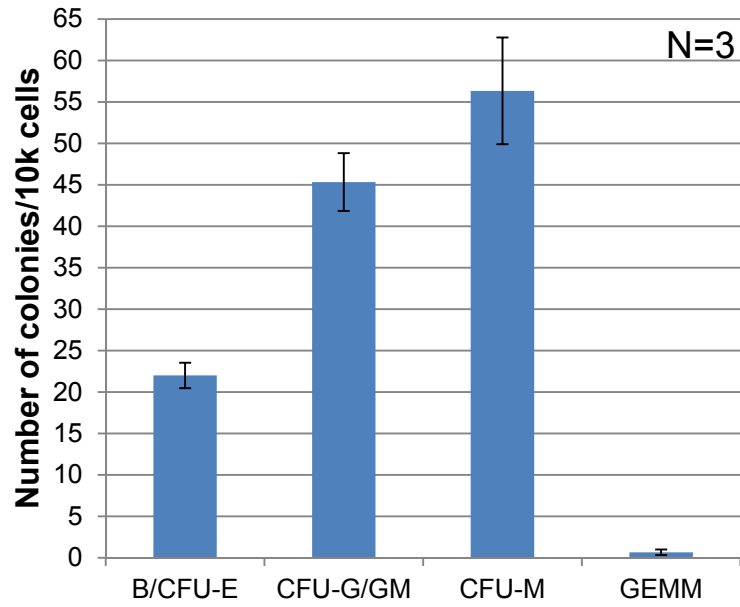


Figure 5-8. FACS analysis of reprogrammed cells revealed a emerged CD34+ and CD45+ population. Under the OP9 stromal cell co-culture condition, a novel CD34+CD45+ cell population showed up about 2 weeks post-infection and continued to expand to day 25 and beyond. Since the retroviral vector expresses 4 iPSC factors as well as GFP, co-expression of GFP and CD34 indicates the 4 iPSC factors are important for the emergence of CD34+ cells.

A Sorting CD34+45+, day 14



B Sorting CD34-45+, day 14

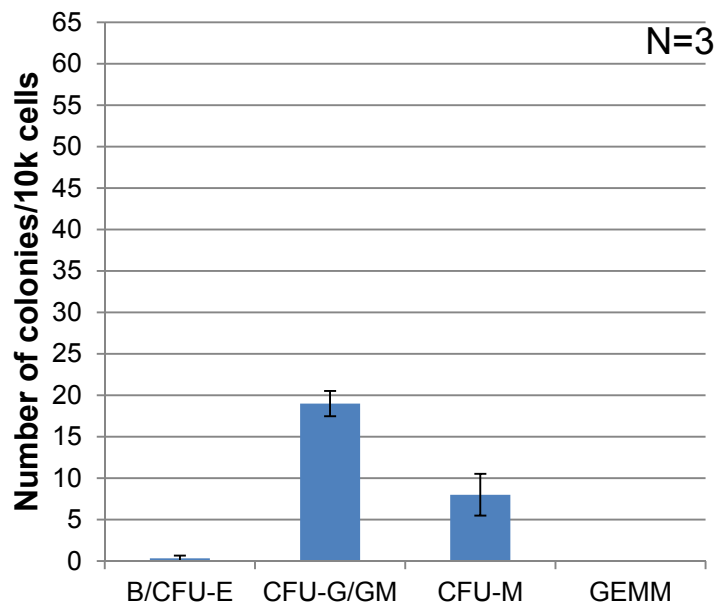


Figure 5-9. Reprogrammed CD34+CD45+ cells showed multi-lineage differentiation potential. (A) Emerged CD34+CD45+ cells were purified and can form various colony-forming progenitor cells in the CFC assay. **(B)** Unexpectedly, CD34-CD45+ cell population can also form colony-forming progenitor cells, although with lower efficiency.

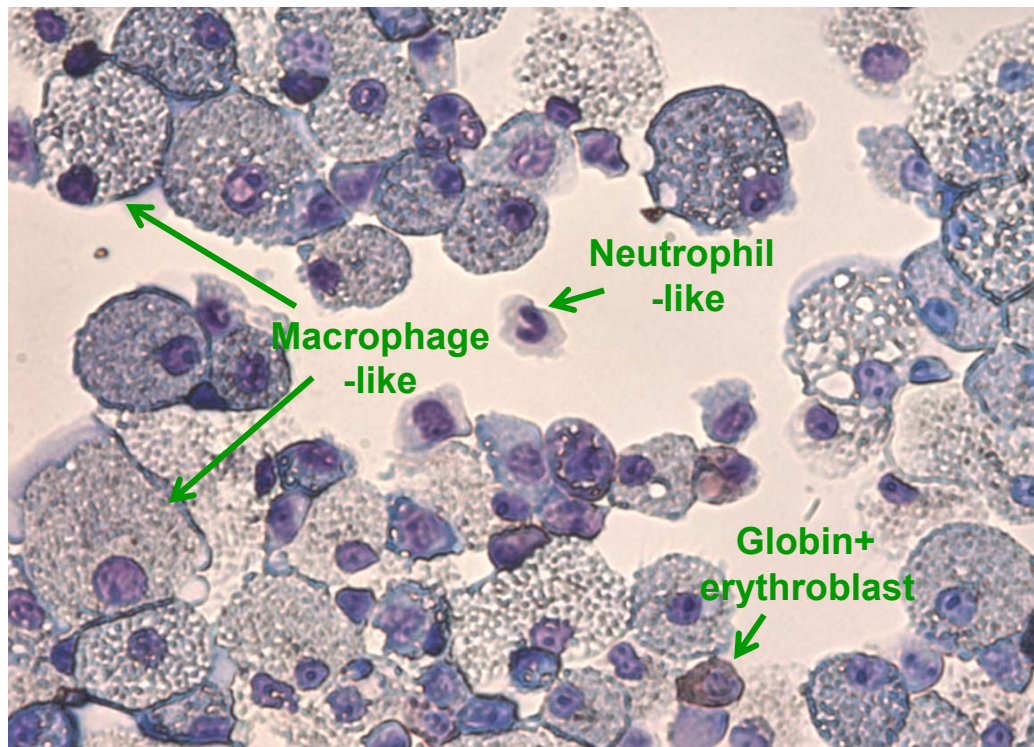


Figure 5-10. Differentiated colony-forming cells harvested from the CFC assay showed various blood cell types in the cytospin assay.

Chapter 6

Conclusions and Future Perspectives

In this thesis work, I tried to reprogram blood cells and address some of key questions surrounding the generation of induced pluripotent stem cells (iPSCs). My findings may help bring us a little closer to the future application of these cells in cell therapy and regenerative medicine.

Reprogramming to pluripotency is a lengthy and stochastic process with extremely low efficiency (Hanna et al., 2009). Identifying an ideal cell type for reprogramming will make iPSC derivation more feasible and significantly enhance the efficiency. Therefore I worked to find a cell type that is amenable for induction within minimal culture period and can be easily acquired. My colleagues in the Cheng lab and I examined the epigenetic profiles of blood cells with age matched fibroblast or ESCs and iPSCs, and found blood cells are closer to ESCs and iPSCs than fibroblasts. In addition, blood can be easily acquired and has become routine practice in hospitals with low risk. Therefore, I targeted blood cells as my starting population for reprogramming.

Most of reprogramming methods depend on viral vectors to delivery transgenes, which makes virus-mediated integration and transgene reactivation risk factors for future usage of iPSCs. To achieve the goal of deriving high quality,

virus-free and integration-free iPSCs, I tested the novel EBNA1/OriP episomal vectors expressing 5-6 reprogramming factors which were solely plasmid-based. I demonstrated CB CD34+ cells can be easily reprogrammed. The EBNA1/OriP episomal vectors also can generate iPSCs from adult CD34+ cells. Inclusion of a second episomal vector Tg had significant effect on refractory cell types.

I also devised a new technique to propagate mononuclear cells (MNCs) from either fetal or adult blood. Instead of expanding MNCs to the lymphoid lineage which may induce DNA rearrangement and lead to altered genome, I used culture conditions for erythroblast proliferation. The expanded cells showed rapid growth and displayed erythroblast phenotypes. Most importantly, these robustly proliferating cells from CB origin were very amenable for episomal vector-mediated reprogramming. The relative efficiency of cells from adult origin (peripheral or bone marrow blood cells) was low, but sufficient iPSC colonies can still be readily generated. Detailed studies like QPCR, Southern blot and whole genome sequencing proved episomal vector-derived iPSCs are footprint-free with minimal mutations. Other complementary studies showed the resulting iPSCs can differentiate to multiple cell lineages. Therefore, my thesis research developed a highly practical method to derive footprint-free human iPSCs from few milliliter blood cells from adult PB MNCs. This method will accelerate the usage of iPSCs in both research and clinical applications.

The ability to reprogram somatic cells from healthy or diseased donors makes it possible to generate iPSCs carrying disease-causing mutations for disease modeling and drug development. To date, many iPSC lines from patients

of neurological, cardiac, blood and other diseases have been reported (Wu and Hochedlinger, 2011; Bellin et al., 2012). Recent studies also suggest that disease-specific iPSCs after differentiating to relevant cell types display certain disease features (Ye et al., 2009; Ebert et al., 2009; Carvajal-Vergar et al., 2010). However, most of disease-specific iPSCs are created by viral vector infection. Viral vector insertion to variant loci can render reprogrammed cells a heterogeneous population. Thus, individual iPSC clones may have subtle differences for differentiation and disease-recapitulation potential. To solve this problem, a reprogramming method that won't cause integration variation is preferred. For this part of study, I used our integration-free technology to derive disease-specific iPSCs from a patient with a HIF2 α mutation. Sequencing data confirmed the resulting iPSC line carried the disease-initiating mutation. It will be interesting to test if this cell line shows any disease related phenotypes after differentiation. Such integration-free iPSC line should provide an unbiased comparison for disease recapitulation. So far most of studies only used rather few iPSC lines to reproduce disease phenotypes. In the future, larger sets of genetically matched iPSCs from one patient or multiple iPSC lines from individual patients are required to verify disease-specific iPSCs can recapitulate disease traits.

I also addressed the question of how to engineer iPSCs to facilitate cell differentiation. Although great efforts have been invested to improve direct PSC differentiation to functionally mature cells of various lineages with capability for therapeutic transplantation, to date the results are largely unsuccessful. This

issue undoubtedly represents one of the major challenges in PSC application. Hematopoietic development is arguably the most characterized cellular differentiation system in mammals. In addition, hematopoietic lineage differentiation is among the easiest to generate from PSCs. Therefore, developing a reporter system capable of lineage tracing will be beneficial. I first transfected iPSC line with *piggyBac* transposon vectors expression GFP. Long term passage of this transgenic cell line affirmed the stable and intensive expression of GFP. This cell line will be appropriate for *in vivo* lineage tracing assay.

Creation of an iPSC line with hematopoietic-specific GFP reporter gene will greatly assists us to trace hematopoietic lineage commitment as well as to study mechanisms governing blood differentiation. To this end, I constructed a *piggyBac* vector carrying fragments of WAS-promoter and RUNX1 enhancer to drive GFP. After transfecting to BC1 iPSC line, I showed undifferentiated iPSCs did not express GFP. Importantly, I found some CD45⁺ hematopoietic cells differentiated from this reporter cell line expressed GFP. Although not all CD45⁺ cells co-expressed GFP gene, almost every GFP⁺ cells acquires CD45⁺ signal. At this time point, it is unclear why the GFP expression is not homogeneous in CD45⁺ population. One possibility is that CD45⁺ cells are heterogeneous and only partial of them were regulated by WAS promoter and/or RUNX1 enhancer to display GFP expression. It would be interesting to examine the differentiation potential of GFP ⁺/₋ population. In sum, I have shown the usefulness of WAS promoter/RUNX1 enhancer to specifically express GFP in iPSC-derived

hematopoietic cells. This reporter iPSC line may serve as an ideal tool to isolate early hematopoietic progenitor cells and screen molecules that can enhance hematopoiesis.

In the last part of this thesis, I sought to reprogram blood cells to cells relevant for clinical application. Although iPSCs provide a potential platform for cell therapy, iPSC *per se* is not clinically relevant. A pretreatment is required to differentiate iPSCs into therapeutically meaningful cells before we transplant cells to patients. But current differentiation protocol can't derive any therapeutic cells from iPSCs.

Recent studies showed the standard Yamanaka four factors were used to convert cell somatic cells to multiple lineages (Efe et al., 2011; Kim et al., 2011; Szabo et al., 2010). The assumption of this approach is that the four factors may “unlock” original cell fate and bring cells to a temporally plastic state. Different developmental cues and culture conditions would coax cells to the desired lineage. Therefore, I asked if four iPSC-factors can not only reprogram blood cells to iPSCs but also convert to therapeutic useful cell types such as hematopoietic stem/progenitor cells (HSPCs). I used the culture condition to expand blood mononuclear cells (MNCs) to erythroblast lineage. The expanded cells remained pan-hematopoietic marker CD45 positive and exhibited erythroblast markers, demonstrating they were lineage committed blood cells. To have a defined starting cell population, I sorted cells displaying erythroblast marker CD36 and excluded common HSPC marker CD34.

After retroviral vector transfection, I cultured cells under conditions favoring HSPC expansion. I tested both OP9-free and OP9 stromal cell co-culture conditions. Strikingly, transfected cells acquired CD34 marker 14 days after transfection. The coincident expression of CD34 and CD45 markers implied some cells may gain HSPCs traits. I next collected CD34 and CD45 double positive cells for CFC assay and found they can form colony forming units of variant types, an evidence of multi-lineage differentiation potential. Taken together, my data showed it is possible to directly reprogram lineage committed blood cells to progenitors using the four iPSC factors. This phenomenon is analogous to dedifferentiation in Waddington's epigenetic landscape.

My results seemed promising for the proof-of-principle of dedifferentiation, but it also raises various questions. Is there HSPC contamination in the starting cell population? Although sorting CD36⁺CD34⁻CD45⁺ triple positive cells diminished the possibility of presumable HSPC contamination, clonal isolation of starting cells should be tested to rule out the possibility. Second, to define more rigorously the phenomenon of dedifferentiation, we need better prove iPSC generation is not involved in the process. Although the culture condition favors the HSPC expansion, it is also permissive for potential iPSC generation and generation of CD34⁺ cells from iPSCs. The time lapse assay may help to figure out if iPSC is one of intermediate during CD34⁻ to CD34⁺ transition. Third, capability of long term culture and expansion of CD34⁺CD45⁺ remains a huge challenge. The limitation of keeping reprogrammed CD34⁺CD45⁺ cells in culture may hamper the further characterization. Lastly, it is crucial to do functional

characterization of these reprogrammed CD34+CD45+ cells. They could be just artificial product during cell culture without any biological meanings. On the other hand, it is also possible that the cellular reprogramming-mediated dedifferentiation could recapitulate the very early stage of hematopoiesis. And such CD34+CD45+ cells could highly resemble authentic HSPC and hold potential in clinical application. Thus, the global gene expression study and *in vivo* transplantation assay will help us to answer these questions.

Future Perspectives

The groundbreaking discovery of induced pluripotent stem cells (iPSCs) opens up a new era for future personalized cell replacement therapy (**Figure 6-1**). My studies tried to solve several bottlenecks. First, I created an efficient and safe method to generate footprint-free iPSCs from easily accessible blood, including neonatal and adult origins. I also demonstrated the feasibility of generating patient-specific iPSCs which can serve as an ideal model for disease modeling and drug screening. To extend to usage of iPSCs, I developed a reliable hematopoietic reporter system. Hopefully it can help us to develop a better culture technique to perform direct hematopoietic differentiation from PSCs. In addition to standard iPSC-involved route for cell replacement therapy, I created a new path that is to directly reprogram lineage committed blood cells to potential clinically useful cell types—HSPCs (**Figure 6-1**). If a patient's disease does not result from genetic mutation and no need of gene correction, one day we may go through this shorten path and bypass the route of iPSC generation

and gene correction. My data could open a regime of dedifferentiation for future regenerative medicine application. In conclusion, the field of stem cell biology and regenerative medicine is rapidly growing, hopefully one day my research works will be translated into the clinic to cure diseases and save lives.

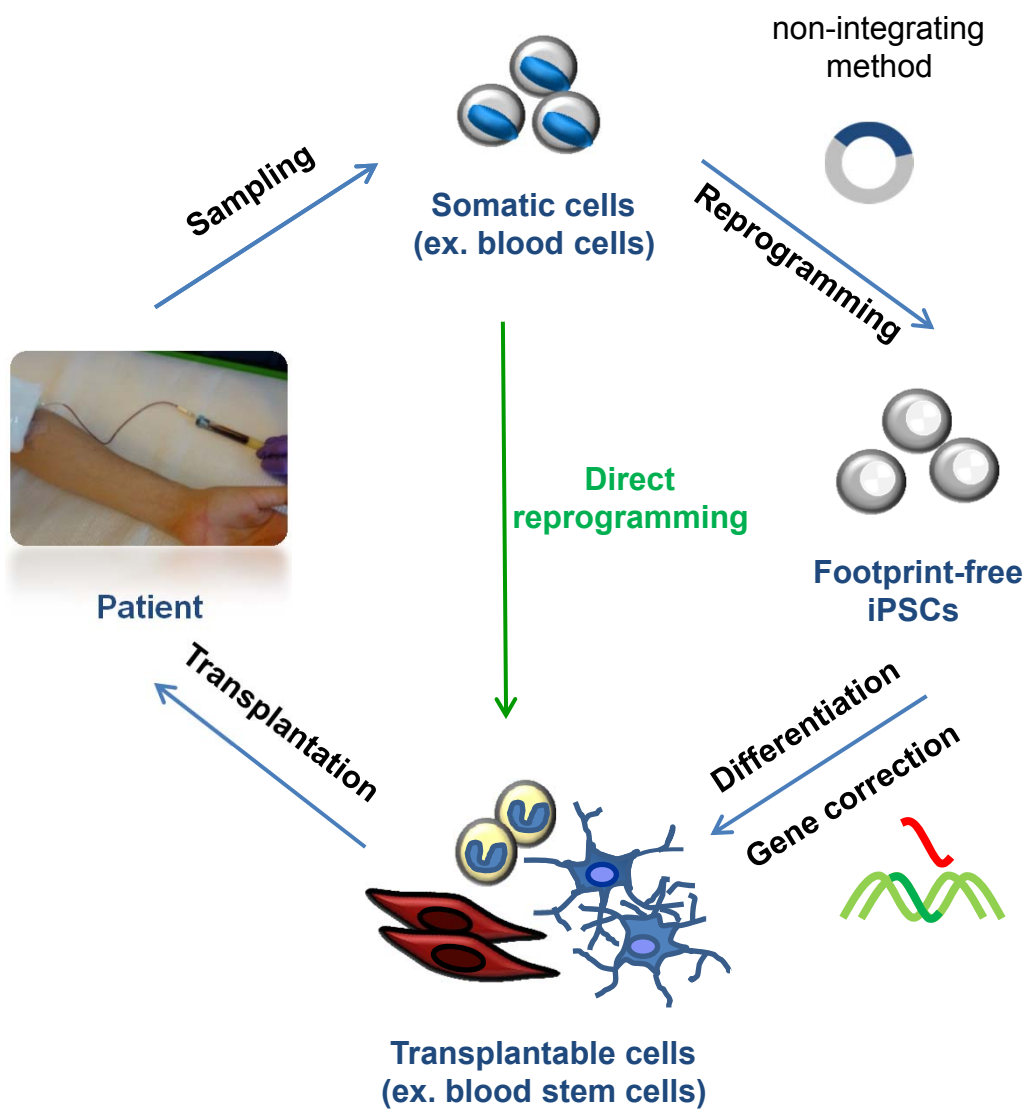


Figure 6-1. The paradigm of reprogramming cells for regenerative medicine. The paradigm begins with collecting patients' somatic cells (such as blood or skin cells). If a patient's cells contain disease-causing genetic mutations, cells will be reprogrammed to iPSCs using non-integrating reprogramming technique and the genetic mutations will be corrected in the iPSC stage. Next, immune-matched disease-free iPSCs will be differentiated to functional and transplantable cell types for autologous cell therapy. In the situation where gene correction is not required, the patient's cells can be directly reprogrammed to clinically useful and transplantable cell types for subsequent cell therapy and bypass the generation of iPSCs.

Materials and Methods

Approvals of Using Human ESCs, iPSCs and Primary Cells from Anonymous Donors

The experimental designs using human ESCs and iPSCs were approved by the Institutional Stem Cell Research Oversight (ISCRO) committee in the Johns Hopkins University. Use of anonymous human samples for laboratory research was approved by the Institutional Review Board.

Constructions of Episomal Vectors

All transgenes were cloned into the EBNA1/OriP-based pCEP4 episomal vector (Invitrogen). A single EBNA1/OriP vector expressing five mouse cDNAs Oct4, Sox2, Klf4, c-Myc and Lin28 that are linked by 2A sequences and under the control of the synthetic CAG promoter was constructed in the pCEP4 plasmid backbone, analogous to a *piggyBac* transposon plasmid we previously used (Mali et al., 2010). The cDNA for the SV40 large T antigen was PCR cloned and inserted into the pCEP4 plasmid backbone as we did previously into a lentiviral vector (Mali et al., 2008). The expression cassette for expressing p53-shRNA was obtained by digesting a fragment from a pSUPER-based vector, and was inserted into the pCEP4 plasmid. The pSUPER-shRNA plasmid was purchased from Addgene (Cambridge, MA). The Thomson/Yu three EBNA1/OriP plasmids used in the combination #6 were also purchased from Addgene and used in transfection at the reported ratio (Yu et al., 2009).

Primary Cells from Anonymous Donors

Most of human primary hematopoietic mononuclear cells (MNCs) obtained from anonymous donors were collected and frozen at AllCells, LLC (Emeryville, CA), including human MNCs expressing a high-level of the CD34 surface marker (CD34+) from newborn cord blood (CB), adult bone marrow (BM) and peripheral blood (PB) after G-CSF mobilization. CD34+ cells were isolated at AllCells or in our lab by the CD34 MACS beads (Miltenyi, Auburn, CA), although CD34 depleted MNCs also contain MNCs expressing a low-level CD34 surface expression. For selected CB and PB samples donated by the parents or patients via their doctors, we isolated MNCs using a standard protocol by Ficoll-Paque Plus (p=1.077) purchased from GE HealthCare Bio-Sciences AB (Uppsala, Sweden).

Culture Media and Conditions for Expanding Human iPSCs, and Karyotyping

The human ESCs and established iPSCs were maintained on a mitotically inactivated mouse embryonic fibroblast (MEF) feeder layer in KNOCKOUT/DMEM medium (Invitrogen) supplemented with 20% KNOCKOUT Serum Replacement (KSR) (Invitrogen), 2 mM L-glutamine (Invitrogen), 2 mM nonessential amino acids (Invitrogen), 1x antibiotic/antimycotic mix (Invitrogen), 0.1 mM β -mercaptoethanol (Sigma), and 10 ng/ml basic fibroblast growth factor (bFGF). All cytokines were purchased from Peprotech (Rocky Hill, NJ) if not otherwise indicated. Karyotyping of human ESCs and iPSCs by G-banding (300-

500 bands) was examined by a certified cyto-geneticist by the method previously described (Mali et al., 2008; Ye et al., 2009; Mali et al., 2010).

Constructions of WAS-Promoter/ RUNX1-Enhancer Reporter Vector

The replication-defective HIV-based lentiviral vector containing 170 bp promoter of WAS with GFP was kindly provided by Dr. Morrone (Leuci et al., 2009). The WAS-GFP fragment was PCR amplified in order to create new restriction enzyme cutting site KpnI at 5' position and cloned to TOPO vector. Then the full WAS-GFP fragment was cut by KpnI/XhoI from TOPO vector and inserted into an empty *piggyBac* vector. The RUNX1 enhance element was obtained by PCR amplification from human iPSCs with new created NheI/KpnI site and cloned to TOPO vector. Then the enhancer element was isolated from TOPO vector by NheI/KpnI and cloned to upstream of WAS-GFP fragment.

Infection of BC1 iPSC Line with Hematopoietic Reporter Vector

2 million BC1 iPSCs were re-suspended in 100 µl mESC nucleofection solution (Amaxa/Lonza) with 5 µg of WAS/RUNX1 reporter vector or 5ug CAG-GFP vector and 5 µg of transposase. We used nucleofector II machine with A-023 program for one-time transfection of BC1 iPSCs. The transfected iPSCs were cultured on MEFs in human ESC medium for 2 days. From day 3, we used 1 µg/ml puromycin to select transfected cells for consecutive 7 days. To isolate cells with high expression of CAG-GFP, growing cells after puromycin selection were subject to cell sorting and reseeded on MEF-coated plates with human ESC

media. Only individual clones stably express GFP were picked and expanded. To derive clonal hematopoietic reporter cell lines, various clones were picked up and subject to spin-EB hematopoietic differentiation to isolate the clone with high percentage of GFP expression.

Human CD34⁺ Cell Culture and Priming for Reprogramming by Episomal Vector Transfection

Frozen human CD34⁺ cells from CB, adult BM and PB after G-CSF mobilization were purchased from AllCells. After thawing, CD34⁺ cells were cultured for 4 to 5 days under a serum-free medium (SFM): containing 50% IMDM with 50% Ham's F12 (Invitrogen), synthetic lipids (Invitrogen), ITS supplement (insulin-transferrin-selenium and 5 mg/ml BSA), 50 ug/ml of ascorbic acid and 2 mM glutamine (all from Sigma). The SFM is further supplemented with cytokines: SCF (100 ng/ml), Flt3 (100 ng/ml), human TPO (20 ng/ml) and IL-3 (10 ng/ml) for expanding CD34⁺ cells, similar to what we previously described (Ye et al., 2009). For reprogramming with EBNA1/OriP-containing episomal vectors, single or combinations of plasmids (up to 10 µg total) were co-transfected into 1-2X10⁶ human CD34⁺ cells via nucleofection (human CD34 cell solution [Nucleofector Kit VPA-1003] with the U-008 program, Amaxa/Lonza). Transfected CD34⁺ cells were cultured in one well of a 12-well plate in the same medium and cytokines for another 2 days. Subsequently, transfected cells were transferred to 3 to 6 wells of MEF-coated 12-well plates and cultured in MEF medium (DMEM + 10% FBS). Plates with seeded cells were spun at 500 rpm for 30 min at room temperature to

help cells attach to MEF-coated plates. The next day (day 3), the MEF medium was replaced with the ESC medium. Small organic molecules that enhance reprogramming such as sodium butyrate (NaB, 0.25 mM) and Valproic Acid (VPA, 0.5 mM) were also added at this time when indicated. The ESC culture medium was exchanged every 3 other day. Starting day 9 post-transfection, MEF conditioned medium (CM) was used to sustain the development of colonies. Colonies with morphology similar to iPSC colonies were readily visible on day 6 to 10 post-transfection of CB CD34⁺ cells, and on day 10 to 14 post-transfection of adult BM and PB CD34⁺ cells.

Human MNC Culture and Priming for Reprogramming by Episomal Vector Transfection

After thawing, frozen MNCs from CB and PB were cultured in the SFM supplemented with the following cytokines and a hormone: SCF (50 ng/ml), IL-3 (10 ng/ml), EPO (2 U/ml, R&D Systems), IGF-1 (40 ng/ml), and dexamethasone (1 μ M, Sigma). The media were replenished at day 3 and 6. By day 8-9 when an overt sign of cell division was observed, cells were harvested for nucleofection as well for various analyses. Approximately 5-fold cells were obtained for CB at day 8 and 1.5-fold for PB MNC at day 9. After optimizing nucleofection of the cultured MNCs to achieve maximal cell survival and gene transfer, we settled down with the human CD34 cell solution but with the T-016 program for one-time transfection of 2×10^6 cultured cells with 10 μ g plasmid DNA. The transfected MNCs were cultured in the MNC medium with the same cytokines for 2 days and

then further reprogrammed on MEF feeder cells as described above with cultured CD34⁺ cells (Ye et al., 2009). The cultured human PB and CB MNCs for 8-9 days were also analyzed by standard hematopoietic assays, in comparisons with newly thawed stock of the same types or after one-day culture. They were analyzed by flow cytometric analysis with a panel of monoclonal antibodies and by colony-forming assays for myeloid- and erythroid progenitor cells (Ye et al., 2009). The stained cells were analyzed using FACS Calibur flow cytometer and CellQuest software (Becton Dickinson Immunocytometry Systems, San Jose, CA) as we described previously (Ye et al., 2009).

TRA-1-60 Live Staining for Identifying Fully Reprogrammed Clones

TRA-1-60 antibody (mouse IgM, used as 1:300 of MAB4360, Millipore) and Alexa555-conjugated antimouse IgM secondary antibody (1:500, Invitrogen) were diluted in ESC medium and added in combination into reprogramming cell culture plates. The cell culture plates with sizable ESC-like colonies were incubated in 37°C for 1 hour before medium was changed to fresh ESC medium or CM after one time wash. TRA-1-60 positive colonies were identified under an inverted fluorescence microscope, and can be picked up (within 1 day after staining) for further expansion and characterization. For quantifying reprogramming efficiency we counted numbers of TRA-1-60⁺, hESC-like colonies as well as TRA-1-60⁻ colonies. The efficiency is plotted as numbers (mean +/- SEM) of TRA-1-60⁺ cells per unit of nucleofected cells after culture

and priming (10^6 for CD34+ cells and 2×10^6 for MNCs). When numbers of duplicates are sufficient ($n \geq 9$), we performed t-test for paired samples assuming a normal distribution and equal variances. P value of ≤ 0.05 is considered to be significant.

Immuno-Staining of Undifferentiated iPSCs

To characterize clones by immuno-staining of various markers, cultured human iPSCs were fixed by 4% paraformaldehyde in PBS for 15 min, followed by permeabilization (required for detecting intracellular antigens) by 0.1% Triton X-100 for 15 min, then washed with PBS for 3 times. The fixed samples were incubated with the following primary antibodies for 2 hours at room temperature: anti-TRA-1-60 (1:300), anti-SSEA-4 (1:10, MC-813-70, mouse IgG, Developmental Studies Hybridoma Bank, Iowa City, IA), anti-NANOG (1:100, 1 μ g/ml, rabbit IgG, Peprotech), anti-OCT4 (1:100, sc-5279, mouse IgG, Santa Cruz Biotechnology). After a brief wash with PBS, Alexa555 conjugated goat anti-rabbit or anti-mouse secondary antibodies (1:500, Invitrogen) were used for one-hour incubation to visualize the cells together with DAPI nuclear staining. In order to visualize the alkaline phosphatase (AP) activity, iPSCs were fixed and stained with Sigma's FAST BCIP/NBT (B5655, Sigma). The AP assay was performed as per the manufacturer instructions.

Embryoid Body (EB) Formation, Differentiation and Immuno-staining

EBs were formed using differentiation medium similar to culture medium except withdrawing bFGF and replacing 20% KSR with 20% FBS. Human iPSCs (near confluent) from 2 wells of a 6-well plate were used for EB formation and cultured in one well of ultra-low attachment 6-well plate. After 8 days, EBs were transferred to gelatin-coated 24-well plates for additional 2-4 days of attachment. Immuno-staining of EBs is similar to that of human ESCs or conventional iPSCs, with the following primary antibodies used: mouse anti- β III-tubulin (1:1000, Sigma), mouse anti-actin, smooth muscle isoform (1:500, CBL171, Millipore), rabbit anti-AFP (1:500, DAKO). Alexa555-conjugated goat anti-mouse or anti-rabbit secondary antibodies (1:500, Invitrogen) were used for visualization. The slides were last counterstained with DAPI to show cell nuclei.

Teratoma Formation Assay of Pluripotency

The use of immuno-deficient mice for the teratoma formation assay was approved by the Animal Care and Use Committee at Johns Hopkins University. Approximately five million iPSCs were harvested by Collagenase IV (Sigma) digestion (from one plates of 6-well plates), washed with PBS and resuspended in 200 μ L diluted (1:1) Matrigel solution (BD). Cells were injected intra-muscularly into *Rag1*^{-/-}*γc*^{-/-} mice or NOD/SCID/IL2RG(γ c)^{-/-} immuno-deficient mice with a further reduced level of natural killer cells. Teratomas were excised 5-10 weeks after injection (Ye et al., 2009; Mali et al., 2010; Mali et al., 2008). After sectioning, slides containing various regions of teratomas were stained by H&E.

Complex structures with various cell types were examined at both low and high magnitude.

Genome-Wide SNP Assay to Assess Genome Integrity

The Illumina's Omni1_Quad BeadArray chip (Illumina) containing $>10^6$ informative SNPs was used. The array analysis was performed by the Johns Hopkins SNP Center as part of Center for Inherited Disease Research (CIDR, www.cidr.org). Based on 1,140,419 SNPs identified and the results with 2 control (CIDR11993 and CIDR10860) genomic DNA samples that have been previously sequenced and included in each run, the Johns Hopkins SNP Center reported a 0.27% genotyping error rate in this run, within the normal range. Genomic DNA (~2 µg) isolated by the Qiagen DNAeasy kit from CT5 and CTN4 iPSCs (p7) that are derived from CB CD34+ cells, and the CD34- cells of the original CB donor (AllCells, #100204) were analyzed. The dis-concordance rate between CT5 or CTN4 iPSCs and the CD34- CB cells is 0.02% or 0.04%, which is much lower than the basal error rate level. In comparison, the dis-concordance rates between the CD34- cells (or iPSCs) and unrelated CIDR11993 or CIDR10860 controls are ~40%. Therefore, our data analysis suggest that CT5 or CTN4 iPSC genome DNA is essentially identical to the CD34- cells and no change is found based on the Omni1 SNP analysis.

PCR and Southern Blot Analysis for the Presence of Vector DNA in the Derived iPSCs

Total DNA (genomic and episomal) from cell extracts was isolated using the DNeasy® kit from Qiagen. Equal amounts of DNA (100 ng) isolated from naive cells (before nucleofection) were used as negative control while the transfected sample (harvested at day 2 post-transfection) were used a positive control for vector DNA detection by the 35-cycle PCR reaction. Three sets of primers used for detecting vector DNA (in either episomal or integrated form) are: EBNA1_D: 5'-TTTAATACGATTGAGGGCGTCT-3', EBNA1_U: 5'GGTTTTGAAGGATGCGA TTAAG-3', Tg_F: 5'-GCCAGGTGGGTAAAGGAGC-3', Tg_R: 5'-GGTACTTATA GTGGCTGGGCTGT-3'; SK_F: CCATTAACGGCACACTGCCCCTGT and SK_R: AGGACGGGAGCAGAGCGTCGCTGA (Yusa et al. 2009). The latter two primers are located at the junction the mouse Sox2 and Klf4 genes in the C5 reprogramming gene cassette, and unable to amplify human genomic DNA. Plasmid DNA, either pEB-Tg or pEB-C5, was used as a control template when appropriate, and diluted to amounts equivalent to 1, 0.2 or 0.04 copies per genome of 100 ng genomic DNA from naïve cells. The detection of the single-copy cellular gene beta-actin in the genomic DNA was employed to normalize the amount of genomic or total DNA used for PCR. We also used Southern blot to detect vector DNA in isolated total DNA using a 445-bp EBNA1 specific probe. The probe was synthesized by PCR amplification using primers EBNA1-5': CCCAGGAGTCCCAGTAGTCAGTC and EBNA1-3': GGAATAGCA AGGGCAGTTCCTCG and the DIG-dUTP labeling kit (Roche) as previously described (Zou et al., 2009). 5µg of total DNA from controls and iPSC samples was digested by BamH I. The pEB-C5 DNA plasmid was also digested by BamH

I (single cut) and used as a positive control for detection of unique vector DNA sequence. Plasmid control DNA was diluted in amounts equivalent to 0.8, 8 and 80 copies per genome in 5 µg genomic DNA used. For genomic DNA quality control of iPSC samples, we used an 848-bp *HBB* probe made by PCR using primers 5'-GACTGAGAAGAATTTGAAAGGCG and 5'-TCATCAATTCTGCCATAAATGG. The wild-type genomic DNA after Pme I/EcoR V digestion showed a single 4.3-kb band detected by the *HBB* probe, as expected. Standard Southern blotting and chemi-luminescence detection with CSPD were performed following the instruction manuals of DIG High Prime DNA Labeling and Detection Starter Kit II (Roche).

PCR Analysis to Confirm the Authentic Origins of iPSC Lines (DNA Fingerprinting)

Genomic DNA isolated from various iPSC lines and their parental (or related) somatic cells types from the same donor was used for the PCR-based DNA fingerprinting assay. Two sets of primers purchased from Invitrogen) were used. Details have been described in our previous publication (Ye et al., 2009; Mali et al., 2010; Mali et al., 2008).

Preparation and Concentration of Retro-Viruses for Direct Reprogramming

Coating 15 cm dishes with 12 ml of 50 µg/ml poly-D-lysine (invitrogen) dissolved in PBS for a period of 1 hour. Wash twice with PBS and then dispense 8-10 million 293T cells in standard DMEM (high glucose) + 10% FBS to a total volume

of 20 ml. The next day, add 36 μ l of Lipofectamine (Invitrogen) to 1.2 ml of OPTI-MEM-I and incubate for 5 min at room temperature in a 5 ml polystyrene tube. In parallel, mix 24 μ g of total DNA (i.e. 3 μ g of VSV.G, 6 μ g of retro-gag/pol, and 15 μ g of retroviral vector) into 1.2 ml of OPTI-MEM-I in another 5 ml polystyrene tube. After 5 minutes, mix the DNA mixture with Lipofectamine and incubate for 20 minutes at room temperature. In the meantime, changed to fresh media in the 293T cultures using DMEM (with 1% FBS) to a total volume of 20 ml. Finally, add the DNA-lipid complexes drop wise onto the 293T cultures. Gently swirl the plates and shake back and forth and sideways to mix uniformly and place dishes in the incubator. After 48 and 72 hours, harvested the supernatant from the plates and added fresh DMEM with 1% FBS to a total volume of 20 ml to the plates for another harvest later. To eliminate the 293 cell contamination, filter the supernatant using a 0.45- μ m filter unit (sigma, Whatman). To concentrate viruses by 50-100 folds the Centricon (Plus-20, 20 ml from Millipore) with a cut-off 100,000 NMWL is our method of choice. The filter filtration method also helps to reduce the free form of VSV.G proteins that are toxic to target cells. Fill the device with ~14 ml virus-containing media and spin down for 20 min at 3000 rpm at room temperature. Each device concentrates 14 ml to ~200 μ l each time. Trash the flow through and refill the device with rest virus-containing media until exhausted. Using this procedure starting from a volume of ~40 ml of supernatant we typically concentrate down to about 400 μ l. The final 400 μ l concentrated virus from original 40 ml of virus-containing media is good to infect 600 thousand cells. We recommend the use of freshly made viruses for transduction, but if it is

not possible, we make small aliquots and store them at -80°C where it is stable for 3 months.

Virus Infection for Direct Reprogramming

6×10^5 expanded PBMNCs were re-suspended in 400 μl concentrated virus in 15-ml conical tube and spin down for 2-3 hours at 2000 rpm at room temperature (the scale can be enlarged). After spin-infection, transferred onto retronectin (optional, Takara) coated 1 well of 12- or 24-well. Added 100 μl fresh SFM with cytokines and the final concentration should be SCF 100 ng/ml, IL-3 10 ng/ml, IGF-1 40 ng/ml, EPO 2U/ml, transferrin 100ng/ml, dexamethasone $1\mu\text{M}$ and polybrene 6 ng/ml. We define the day of virus infection as day 0. One day later, added half SFM with cytokines to reduce polybrene concentration to 4 $\mu\text{g/ml}$. On day 2, harvested all cells and mixed with 10 ml PBS and spun down for 5 min at 200g at room temperature. Removed supernatant and resuspended cells in SFM or SFM + 10%FBS or OP9 co-culture medium. From day 2, the cytokines would be SCF 100 ng/ml, IL-3 10 ng/ml Flt3 100 ng/ml, TPO 20 ng/ml, bFGF 10ng/ml. Change media every other day.

Supplemental References

Yusa, K., Rad, R., Takeda, J and Bradley, A. (2009). Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat Methods* 6, 363-369.

Zou, J, Maeder, ML, Mali, P, et al. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* 5, 97-110.

References

- Ajay, S.S., Parker, S.C., Abaan, H.O., Fajardo, K.V., and Margulies, E.H. (2011). Accurate and comprehensive sequencing of personal genomes. *Genome Res* 21, 1498-1505.
- Albihn, A., Johnsen, J.I., and Henriksson, M.A. (2010). MYC in oncogenesis and as a target for cancer therapies. *Adv Cancer Res* 107, 163-224.
- Anokye-Danso, F., Trivedi, C.M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P.J., Epstein, J.A., *et al.* (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* 8, 376-388.
- Bellin, M., Marchetto, M.C., Gage, F.H., and Mummery, C.L. (2012). Induced pluripotent stem cells: the new patient? *Nat Rev Mol Cell Biol* 13, 713-726.
- Ben-David, U., Gan, Q.F., Golan-Lev, T., Arora, P., Yanuka, O., Oren, Y.S., Leikin-Frenkel, A., Graf, M., Garippa, R., Boehringer, M., *et al.* (2013). Selective elimination of human pluripotent stem cells by an oleate synthesis inhibitor discovered in a high-throughput screen. *Cell Stem Cell* 12, 167-179.
- Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., *et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53-59.

Blank, U., Karlsson, G., and Karlsson, S. (2008). Signaling pathways governing stem-cell fate. *Blood* 111, 492-503.

Blyth, K., Cameron, E.R., and Neil, J.C. (2005). The RUNX genes: gain or loss of function in cancer. *Nat Rev Cancer* 5, 376-387.

Bock, T.A. (1997). Assay systems for hematopoietic stem and progenitor cells. *Stem Cells* 15 Suppl 1, 185-195.

Boitano, A.E., Wang, J., Romeo, R., Bouchez, L.C., Parker, A.E., Sutton, S.E., Walker, J.R., Flaveny, C.A., Perdew, G.H., Denison, M.S., *et al.* (2010). Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science* 329, 1345-1348.

Boland, M.J., Hazen, J.L., Nazor, K.L., Rodriguez, A.R., Gifford, W., Martin, G., Kupriyanov, S., and Baldwin, K.K. (2009). Adult mice generated from induced pluripotent stem cells. *Nature* 461, 91-94.

Briggs, R., and King, T.J. (1952). Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proc Natl Acad Sci U S A* 38, 455-463.

Brockes, J.P., and Kumar, A. (2002). Plasticity and reprogramming of differentiated cells in amphibian regeneration. *Nat Rev Mol Cell Biol* 3, 566-574.

Brown, M., Rondon, E., Rajesh, D., Mack, A., Lewis, R., Feng, X., Zitur, L., Learish, R., and Nuwaysir, E. (2010). Derivation of induced pluripotent stem cells from human peripheral blood T lymphocytes. *PLoS One* 5, e11373.

Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. *Nat Rev Genet* 14, 427-439.

Caiazzo, M., Dell'Anno, M.T., Dvoretzkova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T.D., Menegon, A., Roncaglia, P., Colciago, G., *et al.* (2011).

Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* 476, 224-227.

Carvajal-Vergara, X., Sevilla, A., D'Souza, S.L., Ang, Y.S., Schaniel, C., Lee, D.F., Yang, L., Kaplan, A.D., Adler, E.D., Rozov, R., *et al.* (2010). Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature* 465, 808-812.

Cary, L.C., Goebel, M., Corsaro, B.G., Wang, H.G., Rosen, E., and Fraser, M.J. (1989). Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* 172, 156-169.

Chan, E., Ratanasirintrawoot, S., Park, I., Manos, P., Loh, Y., Huo, H., Miller, J., Hartung, O., Rho, J., Ince, T., *et al.* (2009). Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nat Biotechnol* 27, 1033-1037.

Cherry, A.B., and Daley, G.Q. (2013). Reprogrammed cells for disease modeling and regenerative medicine. *Annu Rev Med* 64, 277-290.

Choi, K.D., Vodyanik, M.A., and Slukvin, I.I. (2009). Generation of mature human myelomonocytic cells through expansion and differentiation of pluripotent stem cell-derived lin-CD34+CD43+CD45+ progenitors. *J Clin Invest* 119, 2818-2829.

Chou, B.K., Mali, P., Huang, X., Ye, Z., Dowey, S.N., Resar, L.M., Zou, C., Zhang, Y.A., Tong, J., and Cheng, L. (2011). Efficient human iPS cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. *Cell Res* 21, 518-529.

Cowan, C.A., Atienza, J., Melton, D.A., and Eggan, K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* 309, 1369-1373.

Daley, G.Q. (2003). From embryos to embryoid bodies: generating blood from embryonic stem cells. *Ann N Y Acad Sci* 996, 122-131.

Ding, L., Ellis, M.J., Li, S., Larson, D.E., Chen, K., Wallis, J.W., Harris, C.C., McLellan, M.D., Fulton, R.S., Fulton, L.L., *et al.* (2010). Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* 464, 999-1005.

Doulatov, S., Notta, F., Laurenti, E., and Dick, J.E. (2012). Hematopoiesis: a human perspective. *Cell Stem Cell* 10, 120-136.

Doulatov, S., Vo, L.T., Chou, S.S., Kim, P.G., Arora, N., Li, H., Hadland, B.K., Bernstein, I.D., Collins, J.J., Zon, L.I., *et al.* (2013). Induction of Multipotential

Hematopoietic Progenitors from Human Pluripotent Stem Cells via Respecification of Lineage-Restricted Precursors. *Cell Stem Cell* 13, 459-470.

Ebert, A.D., Yu, J., Rose, F.F., Mattis, V.B., Lorson, C.L., Thomson, J.A., and Svendsen, C.N. (2009). Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* 457, 277-280.

Efe, J.A., Hilcove, S., Kim, J., Zhou, H., Ouyang, K., Wang, G., Chen, J., and Ding, S. (2011). Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nat Cell Biol* 13, 215-222.

Furlow, P.W., Percy, M.J., Sutherland, S., Bierl, C., McMullin, M.F., Master, S.R., Lappin, T.R., and Lee, F.S. (2009). Erythrocytosis-associated HIF-2 α mutations demonstrate a critical role for residues C-terminal to the hydroxylacceptor proline. *J Biol Chem* 284, 9050-9058.

GURDON, J.B., ELSDALE, T.R., and FISCHBERG, M. (1958). Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* 182, 64-65.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creighton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* 462, 595-601.

Hofmeister, C.C., Zhang, J., Knight, K.L., Le, P., and Stiff, P.J. (2007). Ex vivo expansion of umbilical cord blood stem cells for transplantation: growing knowledge from the hematopoietic niche. *Bone Marrow Transplant* 39, 11-23.

Huang, J., Zhang, Y., Bersenev, A., O'Brien, W.T., Tong, W., Emerson, S.G., and Klein, P.S. (2009). Pivotal role for glycogen synthase kinase-3 in hematopoietic stem cell homeostasis in mice. *J Clin Invest* 119, 3519-3529.

Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* 142, 375-386.

Jia, F., Wilson, K., Sun, N., Gupta, D., Huang, M., Li, Z., Panetta, N., Chen, Z., Robbins, R., Kay, M., *et al.* (2010). A nonviral minicircle vector for deriving human iPS cells. *Nat Methods* 7, 197-199.

Kang, L., Wang, J., Zhang, Y., Kou, Z., and Gao, S. (2009). iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell Stem Cell* 5, 135-138.

Kim, D., Kim, C., Moon, J., Chung, Y., Chang, M., Han, B., Ko, S., Yang, E., Cha, K., Lanza, R., *et al.* (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell* 4, 472-476.

Kim, J., Efe, J.A., Zhu, S., Talantova, M., Yuan, X., Wang, S., Lipton, S.A., Zhang, K., and Ding, S. (2011). Direct reprogramming of mouse fibroblasts to neural progenitors. *Proc Natl Acad Sci U S A* 108, 7838-7843.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I., *et al.* (2010). Epigenetic memory in induced pluripotent stem cells. *Nature* 467, 285-290.

Krizhanovsky, V., and Lowe, S.W. (2009). Stem cells: The promises and perils of p53. *Nature* 460, 1085-1086.

Kunisato, A., Wakatsuki, M., Shinba, H., Ota, T., Ishida, I., and Nagao, K. (2010). Direct Generation of Induced Pluripotent Stem Cells from Human Nonmobilized Blood. *Stem Cells Dev.*

Kyba, M., Perlingeiro, R.C., and Daley, G.Q. (2002). HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. *Cell* 109, 29-37.

Ladewig, J., Koch, P., and Brüstle, O. (2013). Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biol* 14, 225-236.

Lee, G.S., Kim, B.S., Sheih, J.H., and Moore, M. (2008). Forced expression of HoxB4 enhances hematopoietic differentiation by human embryonic stem cells. *Mol Cells* 25, 487-493.

Lee, W., Jiang, Z., Liu, J., Haverty, P.M., Guan, Y., Stinson, J., Yue, P., Zhang, Y., Pant, K.P., Bhatt, D., *et al.* (2010). The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 465, 473-477.

Leuci, V., Gammaitoni, L., Capellero, S., Sangiolo, D., Mesuraca, M., Bond, H.M., Migliardi, G., Cammarata, C., Aglietta, M., Morrone, G., *et al.* (2009). Efficient transcriptional targeting of human hematopoietic stem cells and blood cell lineages by lentiviral vectors containing the regulatory element of the Wiskott-Aldrich syndrome gene. *Stem Cells* 27, 2815-2823.

Liang, H.C., Holmes, R., and Zúñiga-Pflücker, J.C. (2013). Directed differentiation of embryonic stem cells to the T-lymphocyte lineage. *Methods Mol Biol* 1029, 119-128.

Lindner, S.E., and Sugden, B. (2007). The plasmid replicon of Epstein-Barr virus: mechanistic insights into efficient, licensed, extrachromosomal replication in human cells. *Plasmid* 58, 1-12.

Liu, Q., Spusta, S.C., Mi, R., Lassiter, R.N., Stark, M.R., Höke, A., Rao, M.S., and Zeng, X. (2012). Human neural crest stem cells derived from human ESCs and induced pluripotent stem cells: induction, maintenance, and differentiation into functional schwann cells. *Stem Cells Transl Med* 1, 266-278.

Loh, Y., Hartung, O., Li, H., Guo, C., Sahalie, J., Manos, P., Urbach, A., Heffner, G., Grskovic, M., Vigneault, F., *et al.* (2010). Reprogramming of T cells from human peripheral blood. *Cell Stem Cell* 7, 15-19.

Lowry, W.E., Richter, L., Yachechko, R., Pyle, A.D., Tchieu, J., Sridharan, R., Clark, A.T., and Plath, K. (2008). Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci U S A* 105, 2883-2888.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., *et al.* (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 1, 55-70.

Mali, P., Chou, B.K., Yen, J., Ye, Z., Zou, J., Dowey, S., Brodsky, R.A., Ohm, J.E., Yu, W., Baylin, S.B., *et al.* (2010). Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells* 28, 713-720.

Mali, P., Ye, Z., Hommond, H.H., Yu, X., Lin, J., Chen, G., Zou, J., and Cheng, L. (2008). Improved efficiency and pace of generating induced pluripotent stem cells from human adult and fetal fibroblasts. *Stem Cells* 26, 1998-2005.

Marchetto, M., Yeo, G., Kainohana, O., Marsala, M., Gage, F., and Muotri, A. (2009). Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS One* 4, e7076.

Marro, S., Pang, Z.P., Yang, N., Tsai, M.C., Qu, K., Chang, H.Y., Südhof, T.C., and Wernig, M. (2011). Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* 9, 374-382.

Massaad, M.J., Ramesh, N., and Geha, R.S. (2013). Wiskott-Aldrich syndrome: a comprehensive review. *Ann N Y Acad Sci* 1285, 26-43.

Matsui, T., Takano, M., Yoshida, K., Ono, S., Fujisaki, C., Matsuzaki, Y., Toyama, Y., Nakamura, M., Okano, H., and Akamatsu, W. (2012). Neural stem cells directly differentiated from partially reprogrammed fibroblasts rapidly acquire gliogenic competency. *Stem Cells* 30, 1109-1119.

Meissner, A., Wernig, M., and Jaenisch, R. (2007). Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol* 25, 1177-1181.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454, 49-55.

Molineux, G., Pojda, Z., and Dexter, T.M. (1990). A comparison of hematopoiesis in normal and splenectomized mice treated with granulocyte colony-stimulating factor. *Blood* 75, 563-569.

Nakano, T., Kodama, H., and Honjo, T. (1994). Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* 265, 1098-1101.

Nanbo, A., Sugden, A., and Sugden, B. (2007). The coupling of synthesis and partitioning of EBV's plasmid replicon is revealed in live cells. *EMBO J* 26, 4252-4262.

Ng, E.S., Davis, R., Stanley, E.G., and Elefanty, A.G. (2008). A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. *Nat Protoc* 3, 768-776.

Ng, E.S., Davis, R.P., Azzola, L., Stanley, E.G., and Elefanty, A.G. (2005). Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. *Blood* 106, 1601-1603.

North, T., Gu, T.L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marín-Padilla, M., and Speck, N.A. (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development* 126, 2563-2575.

Nottingham, W.T., Jarratt, A., Burgess, M., Speck, C.L., Cheng, J.F., Prabhakar, S., Rubin, E.M., Li, P.S., Sloane-Stanley, J., Kong-A-San, J., *et al.* (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood* 110, 4188-4197.

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T., and Yamanaka, S. (2008). Generation of mouse induced pluripotent stem cells without viral vectors. *Science* 322, 949-953.

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Citri, A., Sebastiano, V., Marro, S., Südhof, T.C., *et al.* (2011). Induction of human neuronal cells by defined transcription factors. *Nature* **476**, 220-223.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141-146.

Parolini, O., Berardelli, S., Riedl, E., Bello-Fernandez, C., Strobl, H., Majdic, O., and Knapp, W. (1997). Expression of Wiskott-Aldrich syndrome protein (WASP) gene during hematopoietic differentiation. *Blood* **90**, 70-75.

Percy, M.J., Beer, P.A., Campbell, G., Dekker, A.W., Green, A.R., Oscier, D., Rainey, M.G., van Wijk, R., Wood, M., Lappin, T.R., *et al.* (2008a). Novel exon 12 mutations in the HIF2A gene associated with erythrocytosis. *Blood* **111**, 5400-5402.

Percy, M.J., Furlow, P.W., Lucas, G.S., Li, X., Lappin, T.R., McMullin, M.F., and Lee, F.S. (2008b). A gain-of-function mutation in the HIF2A gene in familial erythrocytosis. *N Engl J Med* **358**, 162-168.

Pereira, C.F., Chang, B., Qiu, J., Niu, X., Papatsenko, D., Hendry, C.E., Clark, N.R., Nomura-Kitabayashi, A., Kovacic, J.C., Ma'ayan, A., *et al.* (2013). Induction of a hemogenic program in mouse fibroblasts. *Cell Stem Cell* **13**, 205-218.

Petrella, A., Doti, I., Agosti, V., Giarrusso, P.C., Vitale, D., Bond, H.M., Cuomo, C., Tassone, P., Franco, B., Ballabio, A., *et al.* (1998). A 5' regulatory sequence containing two Ets motifs controls the expression of the Wiskott-Aldrich syndrome protein (WASP) gene in human hematopoietic cells. *Blood* **91**, 4554-4560.

Pfisterer, U., Kirkeby, A., Torper, O., Wood, J., Nelander, J., Dufour, A., Björklund, A., Lindvall, O., Jakobsson, J., and Parmar, M. (2011). Direct conversion of human fibroblasts to dopaminergic neurons. *Proc Natl Acad Sci U S A* **108**, 10343-10348.

Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varella, I., Lin, M.L., Ordóñez, G.R., Bignell, G.R., *et al.* (2010). A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191-196.

Potocnik, A.J., Nielsen, P.J., and Eichmann, K. (1994). In vitro generation of lymphoid precursors from embryonic stem cells. *EMBO J* **13**, 5274-5283.

Qiang, L., Fujita, R., Yamashita, T., Angulo, S., Rhinn, H., Rhee, D., Doege, C., Chau, L., Aubry, L., Vanti, W.B., *et al.* (2011). Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons. *Cell* **146**, 359-371.

Rowland, B.D., and Peeper, D.S. (2006). KLF4, p21 and context-dependent opposing forces in cancer. *Nat Rev Cancer* **6**, 11-23.

Ruiz, S., Panopoulos, A.D., Herrerías, A., Bissig, K.D., Lutz, M., Berggren, W.T., Verma, I.M., and Izpisua Belmonte, J.C. (2011). A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr Biol* 21, 45-52.

Sauvageau, G., Iscove, N.N., and Humphries, R.K. (2004). In vitro and in vivo expansion of hematopoietic stem cells. *Oncogene* 23, 7223-7232.

Seki, T., Yuasa, S., Oda, M., Egashira, T., Yae, K., Kusumoto, D., Nakata, H., Tohyama, S., Hashimoto, H., Kodaira, M., *et al.* (2010). Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. *Cell Stem Cell* 7, 11-14.

Semenza, G.L. (2009). Involvement of oxygen-sensing pathways in physiologic and pathologic erythropoiesis. *Blood* 114, 2015-2019.

Serwold, T., Hochedlinger, K., Inlay, M., Jaenisch, R., and Weissman, I. (2007). Early TCR expression and aberrant T cell development in mice with endogenous prerrearranged T cell receptor genes. *J Immunol* 179, 928-938.

Serwold, T., Hochedlinger, K., Swindle, J., Hedgpeth, J., Jaenisch, R., and Weissman, I.L. (2010). T-cell receptor-driven lymphomagenesis in mice derived from a reprogrammed T cell. *Proc Natl Acad Sci U S A* 107, 18939-18943.

Shaltouki, A., Peng, J., Liu, Q., Rao, M.S., and Zeng, X. (2013). Efficient generation of astrocytes from human pluripotent stem cells in defined conditions. *Stem Cells* 31, 941-952.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* 6, e253.

Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J., and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* 9, 205-218.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. *Cell* 136, 364-377.

Staal, F.J., and Luis, T.C. (2010). Wnt signaling in hematopoiesis: crucial factors for self-renewal, proliferation, and cell fate decisions. *J Cell Biochem* 109, 844-849.

Staerk, J., Dawlaty, M., Gao, Q., Maetzel, D., Hanna, J., Sommer, C., Mostoslavsky, G., and Jaenisch, R. (2010). Reprogramming of human peripheral blood cells to induced pluripotent stem cells. *Cell Stem Cell* 7, 20-24.

Szabo, E., Rampalli, S., Risueño, R.M., Schnerch, A., Mitchell, R., Fiebig-Comyn, A., Levadoux-Martin, M., and Bhatia, M. (2010). Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521-526.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676.

Takizawa, H., Schanz, U., and Manz, M.G. (2011). Ex vivo expansion of hematopoietic stem cells: mission accomplished? *Swiss Med Wkly* **141**, w13316.

Thier, M., Wörsdörfer, P., Lakes, Y.B., Gorris, R., Herms, S., Opitz, T., Seiferling, D., Quandel, T., Hoffmann, P., Nöthen, M.M., *et al.* (2012). Direct conversion of fibroblasts into stably expandable neural stem cells. *Cell Stem Cell* **10**, 473-479.

Tian, Y., Luo, A., Cai, Y., Su, Q., Ding, F., Chen, H., and Liu, Z. (2010). MicroRNA-10b promotes migration and invasion through KLF4 in human esophageal cancer cell lines. *J Biol Chem* **285**, 7986-7994.

Trowbridge, J.J., Xenocostas, A., Moon, R.T., and Bhatia, M. (2006). Glycogen synthase kinase-3 is an in vivo regulator of hematopoietic stem cell repopulation. *Nat Med* **12**, 89-98.

Tsien, R.Y. (1998). The green fluorescent protein. *Annu Rev Biochem* 67, 509-544.

Unger, C., Kärner, E., Treschow, A., Stellan, B., Felldin, U., Concha, H., Wendel, M., Hovatta, O., Aints, A., Ahrlund-Richter, L., *et al.* (2008). Lentiviral-mediated HoxB4 expression in human embryonic stem cells initiates early hematopoiesis in a dose-dependent manner but does not promote myeloid differentiation. *Stem Cells* 26, 2455-2466.

van den Akker, E., Satchwell, T., Pellegrin, S., Daniels, G., and Toye, A. (2010). The majority of the in vitro erythroid expansion potential resides in CD34(-) cells, outweighing the contribution of CD34(+) cells and significantly increasing the erythroblast yield from peripheral blood samples. *Haematologica* 95, 1594-1598.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035-1041.

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A.H., and Speck, N.A. (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc Natl Acad Sci U S A* 93, 3444-3449.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., *et al.* (2010). Highly efficient reprogramming to

pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* 7, 618-630.

Wilber, A., Linehan, J.L., Tian, X., Woll, P.S., Morris, J.K., Belur, L.R., McIvor, R.S., and Kaufman, D.S. (2007). Efficient and stable transgene expression in human embryonic stem cells using transposon-mediated gene transfer. *Stem Cells* 25, 2919-2927.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature* 385, 810-813.

Wilson, M.H., Coates, C.J., and George, A.L. (2007). PiggyBac transposon-mediated gene transfer in human cells. *Mol Ther* 15, 139-145.

Wu, S.C., Meir, Y.J., Coates, C.J., Handler, A.M., Pelczar, P., Moisyadi, S., and Kaminski, J.M. (2006). piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc Natl Acad Sci U S A* 103, 15008-15013.

Wu, S.M., and Hochedlinger, K. (2011). Harnessing the potential of induced pluripotent stem cells for regenerative medicine. *Nat Cell Biol* 13, 497-505.

Yamagata, T., Maki, K., and Mitani, K. (2005). Runx1/AML1 in normal and abnormal hematopoiesis. *Int J Hematol* 82, 1-8.

Yamanaka, S. (2010). Patient-specific pluripotent stem cells become even more accessible. *Cell Stem Cell* 7, 1-2.

Yan, X.Q., Hartley, C., McElroy, P., Chang, A., McCrea, C., and McNiece, I. (1995). Peripheral blood progenitor cells mobilized by recombinant human granulocyte colony-stimulating factor plus recombinant rat stem cell factor contain long-term engrafting cells capable of cellular proliferation for more than two years as shown by serial transplantation in mice. *Blood* 85, 2303-2307.

Yates, J.L., and Guan, N. (1991). Epstein-Barr virus-derived plasmids replicate only once per cell cycle and are not amplified after entry into cells. *J Virol* 65, 483-488.

Ye, Z., Chou, B.K., and Cheng, L. (2012). Promise and challenges of human iPSC-based hematologic disease modeling and treatment. *Int J Hematol* 95, 601-609.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I., and Thomson, J. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. *Science* 324, 797-801.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318, 1917-1920.

Zhang, Z., Gao, Y., Gordon, A., Wang, Z.Z., Qian, Z., and Wu, W.S. (2011). Efficient generation of fully reprogrammed human iPS cells via polycistronic retroviral vector and a new cocktail of chemical compounds. PLoS One 6, e26592.

Zhao, Y., Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., Zhang, Q., Xiang, C., Hou, P., Song, Z., *et al.* (2008). Two supporting factors greatly improve the efficiency of human iPSC generation. Cell Stem Cell 3, 475-479.

Zhou, H., Wu, S., Joo, J., Zhu, S., Han, D., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., *et al.* (2009). Generation of induced pluripotent stem cells using recombinant proteins. Cell Stem Cell 4, 381-384.

CURRICULUM VITAE

Bin-Kuan Chou

11/25/13

Educational History

Ph.D. expected	2013	Graduate Program in Cellular and Molecular Medicine The Johns Hopkins University School of Medicine Advisor: Linzhao Cheng, Ph.D.
B.S.	2003	Physics/Zoology (dual major)

Professional Experience

Research Assistant 2005-2007 Stem Cells Lab, Genomic Research Center,
Academia Sinica, Taiwan

Scholarship

Taiwan Merit Scholarship 2007

Awards

Ruth D. Vogel Travel Award 2012

Sanofi-Cell Research Outstanding Paper Award 2011

Third place in the best poster award 2007

Taiwan Society for Stem Cell Research Annual Conference

Outstanding Undergraduate Research Award 2003

National Science Council, Taiwan

Journal Publications

1. Wang Y, **Chou BK**, Dowey S, He C, Gerecht S, Cheng L. Scalable expansion of human induced pluripotent stem cells in the defined xeno-free E8 medium under adherent and suspension culture conditions. *Stem Cell Res.* 2013 Aug 9;11(3): 1103-1116.
2. **Chou BK**, Ye Z and Cheng L. Generation and homing of iPSC-derived hematopoietic cells in vivo. *Mol Ther.* 2013 Jul; 21(7):1292-3.

3. Yu Y, Liang D, Tian Q, Chen X, Jiang B, **Chou BK**, Hu P, Cheng L, Gao P, Li J and Wang G. Stimulation of somatic cell reprogramming by ERas-Akt FoxO1 signaling axis. *Stem Cells*. 2013; Epub ahead of print.
4. Dowey SN, Huang X, **Chou BK**, Ye Z and Cheng L. Generation of integration-free human induced pluripotent stem cells from postnatal blood mononuclear cells by plasmid vector expression. *Nat. Protocol*. 2012 Nov; 7(11):2013-21. Epub 2012 Oct 18.
5. Ye Z, **Chou BK** and Cheng L. Promise and challenges of human iPSC-based hematologic disease modeling and treatment. *Int. J. Hematol*. 2012 Jun; 95(6): 601-9. Epub 2012 May 23.
6. Cheng L, Hansen NF, Zhao L, Du Y, Zou C, Donovan FX, **Chou BK**, Zhou G, Li S, Dowey SN, Ye Z; NISC Comparative Sequencing Program, Chandrasekharappa SC, Yang H, Mullikin JC and Liu PP. Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell*. 2012 Mar 2; 10(3):337-44.
7. Zou C, **Chou BK**, Dowey SN, Tsang K, Huang X, Liu CF, Smith C, Yen J, Mali P, Zhang YA, Cheng L and Ye Z. Efficient derivation and genetic modifications of human pluripotent stem cells on engineered human feeder cell lines. *Stem Cells Dev*. 2012 Aug 10; 21(12):2298-311. Epub 2012 Feb 15.
8. Zou J, Sweeney CL, **Chou BK**, Choi U, Pan J, Wang H, Dowey SN, Cheng L and Malech HL. Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood*. 2011 May 26; 117(21):5561-72. Epub 2011 Mar 16.
9. **Chou BK**, Mali P, Huang X, Ye Z, Dowey SN, Resar LM, Zou C, Zhang YA, Tong J and Cheng L. Efficient human iPS cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. *Cell Res*. 2011 Mar; 21(3):518-29. Epub 2011 Jan 18.
10. Mali P, Ye Z, **Chou BK**, Yen J and Cheng L. An improved method for generating and identifying human induced pluripotent stem cells. *Methods Mol Biol*. 2010; 636:191-205.

11. Mali P, **Chou BK**, Yen J, Ye Z, Zou J, Dowey S, Brodsky RA, Ohm JE, Yu W, Baylin SB, Yusa K, Bradley A, Meyers DJ, Mukherjee C, Cole PA and Cheng L. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells*. 2010 Apr; 28(4):713-20.
12. Zou J, Maeder ML, Mali P, Pruett-Miller SM, Thibodeau-Beganny S, **Chou BK**, Chen G, Ye Z, Park IH, Daley GQ, Porteus MH, Joung JK and Cheng L. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell*. 2009 Jul 2;5(1):97-110. Epub 2009 Jun 18.

Book Chapter

Zhaohui Ye, **Bin-Kuan Chou** and Linzhao Cheng. Human Induced Pluripotent Stem Cells from Cord and Adult Blood Mononuclear Cells. Cord Blood: Biology, Transplantation, Banking and Regulation. *aabb Press*, 2012.

Poster Presentations

1. **Chou BK**, Huang X, Shah S and Cheng L. Reprogramming human committed erythroblast expanded from blood to multi-potent blood progenitors. Poster presentation. International Society for Stem Cell Research Annual Conference, Boston, 2013.
2. **Chou BK**, Tsang K, Dowey S, Ye Z, Wang Y, Kusuma S, Gerecht S, Hansen N, Zhao L, Mullikin J, Liu P, Du Y, Yang H, Hwang T, Gao Y, Zeng X and Cheng L. Genomic, epigenomic, and functional analyses of human integration-free iPS cells derived from adult blood cells by transient expression of a single plasmid vector. Poster presentation, International Society for Stem Cell Research Roddenberry Meeting on Cellular Reprogramming, San Francisco, 2012.
3. **Chou BK**, Mali P, Ye Z, Dowey SN and Cheng L. Virus-free Reprogramming of Human Postnatal Blood Cells. Poster presentation. International Society for Stem Cell Research Annual Conference, San Francisco, 2010.